Kiinteistöopin ja talousoikeuden julkaisuja
Espoo 2007 A 40

ON THE HEDONIC MODELLING OF LAND PRICES

Marko Hannonen

Dissertation for the degree of Doctor of Technology to be presented with due permission of the Department of Surveying, for public examination and debate in Auditorium M1 at Helsinki University of Technology, Espoo, Finland, on the 8th of June 2007, at 12 o'clock noon.

Helsinki University of Technology
Department of Surveying
Institute of Real Estate Studies

Teknillinen korkeakoulu
Kiinteistöopin laboratorio
PL 1200
02015 TKK
Puh. (09) 451 3871
Telefax (09) 465 077
E-mail: real.estate.studies@hut.fi

ISBN 978-951-22-8763-5 (printed)
ISBN 978-951-22-8764-2 (pdf)
(http://lib.tkk.fi/Diss/2007/isbn9789512287642/)
ISSN 0785-5079

Multiprint Oy/Otamedia
Espoo 2007

Abstract

In this study hedonic modelling methods beyond the ordinary least squares estimator are investigated in explaining and predicting the land prices in the two submarkets (Espoo and Nurmijärvi) of the Finnish land markets. The first paper deals with the estimation of several parametric hedonic models, including dynamic responses, using recursive estimation technique. The second paper examines the applicability of semiparametric structural time series methods to the optimal estimation of spatio-temporal movements of land prices. The third paper focuses on the robust nonparametric estimation using local polynomial modelling approach in explaining and predicting the land prices. The fourth paper investigates flexible wavelet transforms in the estimation of long-run temporal land price movements (cycles and trends). The final fifth paper uses robust parametric estimator, the three-stage MM-estimator, to explicitly address the problem of outlying and influential data points.

The key observation of this study is that there is much scope for methods beyond the ordinary least squares estimator in explaining and predicting the land prices in local markets. This is especially true in the submarket of Espoo, where the use of unconventional methods of the study showed that significant improvements could be achieved in hedonic models' explanatory power and/or predictive validity when the methods of this research are used instead of the orthodox least squares estimator. In the Espoo case structural time series models, local polynomial regression and robust MM-estimation all generated more precise results in terms of post-sample prediction power than the conventional least squares estimator. The empirical experimentation quite strongly indicated that the determination of land prices in the municipality of Nurmijärvi could be best explained by the use of unobserved component models. The flexible local polynomial modelling and three-stage MM-estimation surprisingly added no value in terms of greater post-sample precision in the Nurmijärvi case.

Keywords: hedonic prices, flexible estimation, parametric estimator, determination of land prices, robustness

Tiivistelmä

Tässä tutkimuksessa tarkastellaan sellaisia hedonisia mallintamismenetelmiä, jotka yleistävät tavallisen pienimmän neliösumman mukaista ratkaisua, kun selitetään ja ennustetaan maanhintoja kahdella osamarkkina-alueella (Espoo ja Nurmijärvi) Suomen maamarkkinoilla. Ensimmäinen artikkeli tarkastelee erilaisten parametristen mallien estimointia käyttämällä rekursiivista estimointitekniikkaa. Toinen artikkeli tutkii semiparametristen rakenteellisten aikasarjamallien soveltuvuutta ajallis-paikallisten maanhintavaihteluiden optimaalisessa estimoinnissa. Kolmas artikkeli keskittyy vikasietoiseen ja ei-parametriseen estimointiin käyttämällä paikallisia polynomimalleja, kun selitetään ja ennustetaan maanhintoja. Neljäs artikkeli tutkii joustavia aalloke-muunnoksia pitkän ajanjakson maanhintojen vaihteluiden (syklien ja trendien) estimoinnissa. Viimeinen viides artikkeli käyttää vikasietoista parametrista estimaattoria, kolmivaiheista MM-estimaattoria, vähentämään mallintamisessa ilmenevien poikkeavien ja vaikutusvaltaisten havaintopisteiden negatiivinen vaikutus.

Tutkimuksen avainhavainto on, että tutkimuksessa tarkasteltuja epästandardeja menetelmiä voidaan soveltaa hyvin käytännön ongelmaratkaisutilanteissa, kun selitetään ja ennustetaan maanhintoja paikallisilla markkinoilla. Tämä pätee erityisesti Espoon hinta-aineistolla, jossa epästandardien menetelmien käyttö johti hedonisiin hintamalleihin, jotka omasivat huomattavasti korkeamman selitysvoimakkuuden ja/tai ennustustarkkuuden kuin tavallisen pienimmän neliösumman mukainen ratkaisu. Espoon osamarkkinoiden tapauksessa rakenteelliset aikasarjamallit, vikasietoinen paikallinen regressioanalyysi ja vikasietoinen MM-estimointi tuottivat tarkempia tuloksia kuin perinteinen pienimmän neliösumman mukainen keino, kun estimoitujen mallien hyvyyttä arviointiin ennustustarkkuuden mielessä eri kriteereillä. Empiirinen tutkimus indikoi varsin voimakkaasti, että Nurmijärven osamarkkinoiden tapauksessa maanhinnan muodostus voitiin parhaiten selittää käyttämällä rakenteellisia aikasarjamalleja. Sen sijaan joustavat polynomimallit ja MM-estimointi eivät tuoneet lisäarvoa mallien paremman ennustustarkkuuden valossa Nurmijärven hinta-aineistolla.

Avainsanat: hedoniset hinnat, joustava mallintaminen, parametrinen estimointi, maanhinnan muodostus, vikasietoisuus

Preface

This doctoral thesis was made for a large part in years 2003-2006 in a research project called "Improving the Reliability of Valuation Methods in a Valuation for Compensation", which was financed by the Ministry of Agriculture and Forestry. The research also got funding from the Graduate School of Real Estate, Construction and Planning (KIRSU). I would like to thank my financiers of making this thesis possible.

In addition, my special thanks go to Professor Kauko Viitanen, Surveyor Counsellor Arvo Kokkonen, MSc Petki Lukin, Doctor Tuomo Heinonen, MSc Helka-Marja Kohonen, Senior Researcher Juhani Väänänen and Doctor Ari Hiltunen who all helped me throughout the research project by sharing their extensive knowledge with me.

I extend my gratitude also to the National Land Survey of Finland for giving me the necessary land price data used in the hedonic modelling of this research and my pre-examiners Doctor Seppo Laakso and Professor Antti Kanto for their valuable comments.

My special thanks go to my spouse Marja for encouraging me during the different phases of the research.

Table of contents

1 Introduction

1.1 Context of the Research

This study relates to the hedonic estimation of land prices. Hedonic methods are often considered as necessary tools in more demanding and complex land value applications in order to objectively minimise the systematic valuation error and in order to produce the necessary quality-adjustments validly and reliably that arise from the heterogeneous nature of separate land parcels. Correct measures of hedonic prices and quality-corrected mean land prices are important to all participants in the land market since misperceived land and implicit prices could lead to inadequate allocation and use of land resources. The importance of correct measurements is further emphasised by the fact that land transactions involve substantive financial transfers. For instance, in Finland alone almost 170 M€ are annually used for acquiring undeveloped sites of one-family houses with over 7000 enclosed transactions. However, the use of conventional hedonic models is plagued with some fundamental difficulties imposing serious threats to their empirical adequacy. This study tries to address these difficult hedonic modelling concerns with workable solutions that can lead to improved accuracy.

The research is quantitative in nature, which is typical for most hedonic price studies. Hedonic functions are empirically estimated via different modern techniques and their fit and/or post-sample predictive performance is measured in depth. The data in the empirical study comes from the Finnish land market: specifically from the submarkets of Espoo and Nurmijärvi.

1.2 Aims of the Study

The land market is one of the most important sectors of the whole economy; in particular, land itself forms a necessary factor of production for all human activities (housing, industry, business, etc.). However, systematic research about spatio-temporal land price movements in these markets has typically been scarce, inflexible and insufficient mainly relying on the results of the classical least squares estimator. The main general purpose of this dissertation is to investigate the applicability of hedonic modelling methods beyond the ordinary least squares estimator in explaining and predicting the land prices. The study tries to extend the hedonic modelling methodology due to Hiltunen [2003] in several ways paying increased attention to the concepts such as nonlinearity and nonstationarity, which are principal components of spatial-temporal structure in the land markets.

The specific aims of this study are related to some important and problematic issues in the hedonic modelling of land prices, which are typically in practice solved unsatisfactory. These include[1]:

- temporal variability of land prices;
- spatial variation of land prices;
- model specification dilemma;
- nonlinear nature of relationships;
- outlying and influential observations.

[1] It ought to be mentioned that the control of uncertainty underlying the inference is another major modelling concern that is not, however, addressed in this study.

1.3 Structure of the Dissertation

The dissertation includes five papers. Paper 1 ("On the Recursive Estimation of Hedonic Prices of Land") deals with the evaluation of hedonic model's congruence and whether constant, invariant parameters can be identified. Several different parametric hedonic models are recursively estimated and empirically compared together in order to find the best model that matches the data in most measurable respects. Paper 2 ("An Analysis of Land Prices: A Structural Time-Series Approach") presents a relatively new modelling framework: the structural time series modelling context. Structural time series or unobserved component models combine the essential features of a regression and time series analysis. In this paper two common hedonic models are estimated by using unobserved components for a trend specification and cyclical variation.

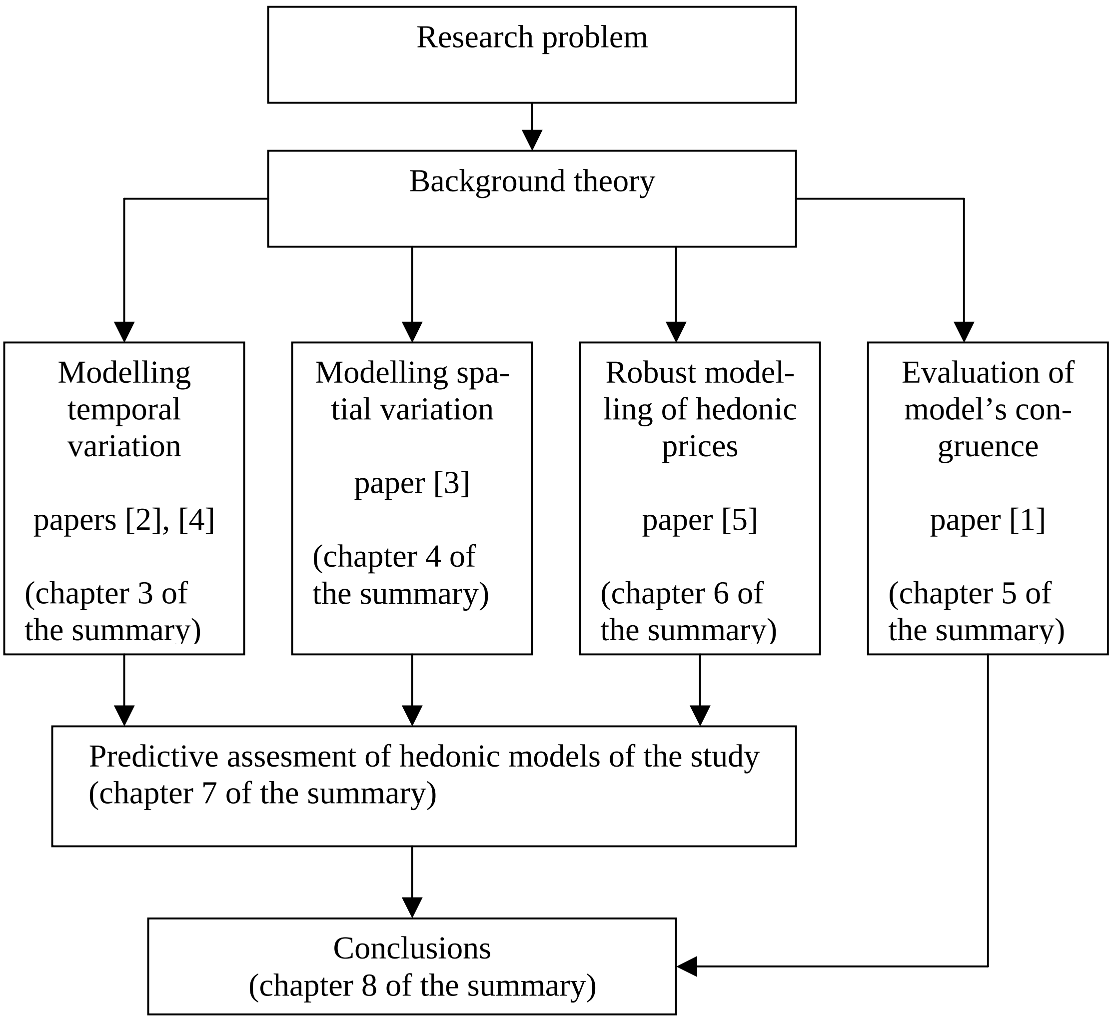

Figure 1. Structure of the study.

Paper 3 ("Forecastability of Land Prices: A Local Modelling Approach") examines the hedonic model specification dilemma by using a nonparametric local regression method, in which the results of the estimation are robustified by a scheme that is a variant of M-estimation. It also gives a proper solution to the spatial variation problem by adjusting empirically to the non-homogeneous features of the data. Paper 4 (“An Analysis of Trends and Cycles of Land Prices Using Wavelet Transforms”) is closely connected to paper 2 as it also considers the optimal estimation of temporal variation of land prices. This paper presents a new approach to temporal variability problem based on wavelet transforms that are nonparametric orthogonal series estimators. This is a highly flexible approach for the determination of components of temporal variation, in which wavelets are capable of providing the necessary time and frequency information simultaneously. Paper 5 (“On the Robust Modelling of Land Prices” (in Finnish)) discusses the use of a robust MM-estimation in the determination of hedonic prices of land. This is a highly fault-tolerant modelling method, which mitigates the negative influence of outlying and influences observations. In figure 1 is depicted the structure of the study.

2 Foundations of Land Markets and Hedonic Theory

2.1 Previous Hedonic Price Studies

The hedonic price studies in land markets have been extensive. In essence, most of the studies are traditional in a sense that global parametric models (usually linear, double-log or semi-log models or some combination of these) are specified in advance by theoretical reasoning or previous knowledge and subsequently they are used to determine the relevant hedonic values. Yet some of the studies apply a more flexible data-driven approach, in which the form of the hedonic function is - at least partly - constructed on the basis of actual observations. Overall the hedonic land price studies indicate that there are several factors that have a major influence on land prices. The most commonly documented factors in these studies include [Hiltunen, 2003, p. 53 & p. 57]:

- parcel size
- permitted building volume
- plot ratio
- distance to the CBD
- selling date
- land type
- population’s income level
- number of inhabitants or its growth
- construction costs
- social status of the area
- distance to the lake and river systems.

Table 1, which is based on the dissertation made by Hiltunen [2003], summarises hedonic price studies that have been carried out in the land markets in Finland with unbuilt sites. The key methodological implication of these studies is that in Finland the research method is solely based on the ordinary least squares usually with double-log model specification. The investigated variable has usually been the unit price of a land parcel; the total sales price has been used rarely. The explanatory power of different models fluctuates quite heavily both inside the same research as submarkets changes and between different studies. The total range of coefficient of determination (in %) is between 19.5 and 99. The length of the time period applied fluctuates also quite heavily; the shortest time period is one year and the longest is 32 years. The number of observations applied in these studies ranges from 589 to 52474.

Table 1. Summary of the hedonic price studies in Finland with unbuilt sites.

Author and date of the study	Data	Method of the analysis	Model specification	Dependent variable	Most significant explanatory variables	Coefficient of determination (%)
Kantola, 1967	589 observations of residential sites with four different municipalities in Finland during the period of 1956-1965	Parametric regression based on ordinary least squares	Double-log model	Unit price	- Distance to the CBD - Plot ratio - Time - Population density	83.1-95.2
Änkö, 1969	1019 observations of residential sites in the neighbouring areas of Helsinki in Finland during the period of 1965-1967	Parametric regression based on ordinary least squares	Double-log model	Unit price	- Size of the community - Distance to the CBD - Intended use of the site - Price differences between population centres	79.8-97.8
Myhrberg, 1971	835 observations of residential sites with six different municipalities in Finland during the period 1935-1967	Parametric regression based on ordinary least squares	Linear model	Unit price	- Distance to the CBD - Parcel size - Percentage of forest in the site	24.7-52.2
Kanerva, 1974	1499 observations of residential and commercial sites with 23 different population centres in the period of 1967-1972	Parametric regression based on ordinary least squares	Double-log model and linear model	Unit price	- Population - Population growth - Income differences - Pleasant environment - Distance to the CBD - Site type - Plot ratio - Distance to the water system	72.2-97.1
Raittinen, 1975	643 observations of sites with three different municipalities in Finland during the period of 1960-1975	Parametric regression based on ordinary least squares	Linear model	Unit price	- Time - Distance to the CBD - Parcel size	34.2-45.5
Lukkarinen, 1981	Sample of land prices with 2260 observantions and 80 different population centres in Finland in year 1980.	Parametric regression based on ordinary least squares	Double-log model	Unit price	- Population - Distance to the CBD - Distance to the water system - Site type - Permitted building volume - Degree of planning - Land policy	74-79
Leväinen & Virkkunen, 1987	10279 observations of residential sites with 84 cities in Finland during the period of 1982-1985	Parametric regression based on ordinary least squares	Linear model	Unit price	- Population - Change of population - Income - Type of the seller	24-94
Lehmusto, Kanto & Sorsa, 1991	Unspecified number of observations of residential sites with three major cities in Finland	Parametric regression based on ordinary least squares	Double-log model	Unit price	- Distance to the CBD - Income level - Population growth - Distance to the	41.2-94.3

	during the period of 1980-1989				sea - Supply of money	
Leväinen, 1991	1537 observati-ons of residential sites with three different cities in Finland during the period of 1982-1989	Parametric re-gression based on ordinary least squares	Double-log model, semi-log model and linear model	Unit price and total price	- Parcel size - Time - Distance to the CBD - Distance to the sea	19.5 – 92.7
Heinonen, 1993	20676 observati-ons in whole Finland on residential sites during the period of 1985-1991	Parametric re-gression based on ordinary least squares	Double-log model	Unit price	- Planning earmark - Permitted building volume - Parcel size - Population - Popultion growth - Type of the seller - Time	24-93
Hiltunen, 1993	Observations of land price index in whole Finland in years 1961-1991	Parametric re-gression based on ordinary least squares	Double-log model	Land price index and its change	- House prices - Interest rate - Income level	40-99
Peltomaa, 2001	52474 observati-ons of residential sites in whole Finland during the period of 1993-1997	Parametric re-gression based on ordinary least squares	Double-log model	Unit price	- Planning status - Permitted building volume - Parcel size - Distance to the water system - Location - Time - Distance to the large-capacity line - Type of the buyer	63.6
Hiltunen, 2003	14732 observati-ons of residential sites with six different cities in Finland during the period of 1985-1998	Parametric re-gression based on ordinary least squares	Double-log model	Total price	- Permitted building volume - Location - Distance to the CBD - Sosio-economic status - Time	52.5-81.1

Table 2 summarises hedonic price studies that has conducted abroad in the land markets with unbuilt or vacant sites[2]. The most common estimation method has been the ordinary least squares usually with double-log, semi-log and linear model specification or some combination of these models; the mixture of double-log and semi-log model is quite typical, which has not been used in the corresponding Finnish studies. Unlike in Finland few nonparametric and semiparametric approaches to the modelling of land prices exist abroad; to be specific in three studies a nonparametric estimator has been used and in one study semiparametric estimator has been applied. Also spatial regression, two-stage least squares and instrumental variable method has been used. The explanatory power of different approaches fluctuates heavily both inside the same research and between different studies. The total range of coefficient of determination (in %) is between 1.1 and 98.8. The length of the time period applied fluctuates also quite heavily; the shortest study period is one year and the longest is 154 years. The number of observations applied in these studies ranges from 50 to 48000.

[2] The summary is not complete as it consists of only those studies that have been presented in the major scientific journals since 1980.

Table 2. Summary of the hedonic price studies abroad with unbuilt sites.

Author and date of the study	Data	Method of the analysis	Model specification	Dependent variable	Most significant explanatory variables	Coefficient of determination
Colwell & Sirmans, 1980	143 observations from two different US submarkets on residential sites from the periods of 1952-1967 and 1977-1978.	Parametric regression based on ordinary least squares	Double-log model and linear model	Unit price and total price	- Time - Distance to the CBD - Parcel size - Type of the seller - Area factor - Corner site	14.7-74.4
Asabere, 1981a	211 observations of sites from Accra, Ghana in years 1974-1978	Parametric regression based on ordinary least squares	Mixture of double-log and semi-model	Total price	- Distance to the CBD - Distance to the sea - Road indicator - Site type - Time - Parcel size	49-65
Asabere, 1981b	120 observations of sites from Kumasi, Ghana in years 1970-1979	Parametric regression based on ordinary least squares	Mixture of double-log and semi-log model	Total price	- Type of the seller - Time - Distance to the CBD - Distance from the university - Nearness of a road - Low-class indicator - High-class indicator - Presence of a river - Parcel size	53
Chicoine, 1981	1400 observations of farmland prices from Illinois, USA in years 1970-1974	Parametric regression based on ordinary least squares	Mixture of double-log and semi-log model	Unit price	- Distance to the CBD - Distance to the nearest freeway exchange - Frontage road type - Neighborhood - Mining / qauarrying land use - Restrictions on land use - Land type - Time - Parcel size - Type of the buyer and seller	52
Bartnett, 1985	2096 observations of residential sites in Perth, Australia during the period of 1977-1978	Parametric regression based on ordinary least squares	Double-log model and linear model	Total price	- Parcel size - Distance to the CBD - Open view - Location in the main road - Distance to the industrial area - Sosio-economic status	66-76
Dunford, Marti & Mittelhammer, 1985	Unspecified number of observations of vacant rural land in Washington, USA, in the year 1978	Parametric regression based on ordinary least squares	Semi-log model	Total price	- Time - Buyer's perception of present development - Future intensity of development - Probability of septic tank installation	63.5

Shonkwiler & Reynolds, 1986	189 observations of vacant rural land in Florida, USA, in years 1973-1981	Parametric regression based on ordinary least squares and instrumental variable estimation	Mixture of double-log and semi-log model	Unit price	- Land type - Parcel size - Proportion of cultivated land - Time - Two local distance measures	70.5
Johnson & Ragas, 1987	110 observations of vacant commercial land in New Orleans, USA, in the period of 1972-1983	Parametric regression based on ordinary least squares	Double-log model, linear model and interactive quadratic model	Unit price	- Time - Distance to the CBD - Location along major streets - Zone type - Distance to the Louisiana super-Dome - Two other local distance measures	52-60
Manning, 1988	Median price data from 94 different cities in USA in year 1980	Parametric regression based on ordinary least squares	Linear model	Unit median price	- Household's income - Population density - Population growth - Climate - Construction costs	74.7
La Croix & Rose, 1989	Observations from 41 different states in USA in year 1980	Parametric regression based on ordinary least squares	Double-log model	Unit price	- Population - Income differences - Population growth - Amount of foreign real estate invest ments - Supply of land	54.0-58.8
Colwell, 1990	200 observations of vacant sites in Illinois, USA, in the period of 1968-1978	Parametric regression based on ordinary least squares	Mixture of double-log and semi-log model	Total price	- Parcel size - Distance to the tower - Distance to the electric transmission line - Four improvement variables - Deck/porch indicator - Holiday Hills/ Windsor Village indicator - Time	77.1-77.2
Asabere, 1991	100 observations of residential, commercial and industrial sites in Philadelphia	Parametric regression based on ordinary least squares	Mixture of double-log and semi-log model	Total price	- Parcel size - Time - Median income - Percentage of abandoned houses - Land type - Local distance measure	40.3
Asabere & Huffman, 1991	100 observations of vacant land sales in Philadelphia in the years 1987-1989	Parametric regression based on ordinary least squares	Semi-log model	Total price	- Parcel size - Time - Income - Distance to the CBD - Historic district location indicator - Percentage of houses boarded up	40-42

McMillen, 1996	1100 observations from Chicago on sites between 1836-1990	Parametric regression based on ordinary least squares and nonparametric regression based on weighted least squares	Global monocentric model, global spatial expansion model and local spatial expansion model	Total price	- Distance to the CBD - Distance to Michigan lake - Distance to the O'Hare airport - Distance to the Midway airport	23.6-96.8
Potepan, 1996	106 observations of vacant sites with different US cities from years 1975, 1980, 1983 and 1985.	Parametric regression based on ordinary least squares and two-stage least squares	Double-log model	Unit price	Two-stage least squares: - House prices - Construction costs Ordinary least squares: - Income level - Population - Service level - Real estate tax - Construction costs - Restrictions on land use	59-71
Colwell & Munneke, 1997	1110 observations of residential and commercial sites from Chicago in the period of 1986-1992	Parametric regression based on ordinary least squares	Double-log model	Unit price and total price	- Distance to the CBD - Parcel size - Distance to Michigan lake	47.7-61.1
Isakson, 1997	363 observations of vacant sites in Denver, Colorado, USA, in the period of 1985-1992	Parametric regression based on ordinary least squares	Mixture of double-log and semi-log model	Total price	- Parcel size - Time - Distance to the CBD - Zoning situation - Type of the buyer and seller	43.3-55.4
Atack & Margo, 1998	Unspecified number of observations of vacant sites from the period of 1835-1900 in New York, USA	Parametric regression based on ordinary least squares	Semi-log model	Unit price	- Distance to the CBD - Corner lot indicator	1.1-71.7
Colwell, 1998	Observations from Olcott's land values blue book in Chicago	Nonparametric regression based on the spline approximation	Piecewise parabolic model	Unit price	- Barycentric coordinates	63.9
Craig, Palmquist & Weiss, 1998	3390 observations of sites from the years 1850 and 1860 in antebellum, USA	Parametric regression based on ordinary least squares	Semi-log model (in levels and first differences)	Unit price	- Canal indicator - Ocean indicator - Lake indicator - Railroad indicator - Time - Ratio of improved acres to total acres - Ratio of county's population to square miles - Four other indicators decribing county's size and location	72.5-78.6
McDonald & McMillen, 1998	943 observations of sites in Chicago from the year 1921	Parametric regression based on the maximum likehood method	Semi-log model with cubic CBD distance variable and CBD distance interacts with other distance measures	Total price	- Distance to the CBD - Distance to the lake - Distance to the el station - Distance to the commuter rail station - Three additional block distance measures	NA

Thorsnes & Mcmillen, 1998	158 observations of undeveloped sites from Portland, Oregon, USA, in the period of 1980-1987	Parametric regression based on ordinary least squares and semiparametric regression based on the Nadaraya-Watson kernel estimator	Mixture of double-log and semi-log model	Total price	- Distance to the CBD - Distance from the freeway interchange - Distance from the arterial street - Time - Median income - Proportion of white people - Median age	78.9-81.1
Colwell & Munneke, 1999	2678 observations of vacant residential, commercial and industrial parcels in Cook County, Illinois, USA in years 1984-1993	Parametric regression based on ordinary least squares	Mixture of double-log and semi-model	Total price	- Distance to the CBD - Parcel size - Distance to Michigan lake - Distance to O'Hare airport - Three other location measures	47-51
Lin & Evans, 2000	50 observations of vacant sites in Taipei city, Taiwan in years 1996 and 1997	Parametric regression based on ordinary least squares	Double-log model and linear model	Total price and unit price	- Parcel size - Type of the buyer - Time -Nearness of a major road	63.8-98.8
Yamazaki, 2000	4368 observations of sites from Tokyo in years 1985-2000	Parametric regression based on ordinary least squares	Double-log model and semi-log model	Unit price and total price	- Parcel size - District - Effect of streets - Distance to the train station	71-90
Miyamoto, Paez & Uchida, 2001	479 observations of sites in the city of Sendain in Japan in the year 1996	Spatial regression	Spatially switching autoregressive model	Unit price	- Distance to the CBD - Distance to nearby malls - Distance to the metro and the train station - Intensity of commercial land	NA
Colwell & Munneke, 2003	1194 observations of vacant land in Chicago metropolitan area, USA in years 1995-1997	Nonparametric regression based on the spline approximation	Piecewise parabolic model	Total price	- Barycentric coordinates - Parcel size - Time	51
Guntermann & Thomas, 2005	48 observations of vacant commercial sites in Scottdale, USA	Parametric regression based on ordinary least squares	Mixture of double-log and semi-log model	Total price	- Parcel size - Distance to the CBD	93.3
Tsoodle, Golden & Featherstone, 2006	48 000 observations of agricultural land for 104 counties in Kansas, USA in years 1986-2000	Parametric regression based on ordinary least squares	Mixture of double-log model and semi-log model	Unit price	- Parcel size - Average annual rainfall - Average productivity - Percentage of irrigated acres - Percentage of improved range land - Population - Road quality - Average agricultural land rental price - Four transactive factors	70.7

Hedonic approaches contain some serious problems. The main criticism against hedonic price method relates to the conventional statistical problems that are encountered in a regression context [Wallace, 1996]. A first complaint is that the proper identification of characteristics vector is not always possible resulting to a model specification bias, e.g. some of the attributes may be latent variables and a valid proxy may be missing or hard to come by. This problem exists both in the traditional parametric approach based on the ordinary least squares and in more sophisticated approaches, such as a nonparametric or a semiparametric approach. A second complaint concerns the assumption of the correct functional form that must be imposed for estimation purposes. This problem is not as severe in flexible nonparametric and semiparametric approaches as in the common parametric approach because nonparametric and semiparametric approaches can learn many of the systematic features between the dependent variable and a set of attributes from the available data. A final complaint relates to the sample selection bias, i.e. the data rarely constitutes a sample that could be termed as random. Typically a judgement sample is analysed, which is a real threat to a valid statistical inference in parametric, semiparametric and nonparametric approaches.

2.2 Nature of Land Markets

Land, in an economic sense, is defined as the entire material universe outside of people themselves and the products of people. It includes all natural resources, materials, airwaves, as well as the ground. A market is a set of arrangements which allows people to trade, normally governed by the theory of supply and demand, so allocating resources through a price mechanism and bid and ask matching so that those willing to pay a price for something meet those willing to sell for it. In particular, land markets perform four different and important functions [Dowall, 1993]: (1) they bring buyers and sellers together to facilitate transactions; (2) they set prices for land parcels; (3) they allocate land by setting prices so that the land market "clears", i.e., the quantity of land offered for sale equals the quantity of land demanded; and (4) land prices play an important role in ensuring land is efficiently used.

Land markets have some distinguishing features [Hiltunen, 2003, p. 7] that separate them from other commodity markets. Land itself is the most basic resource and a source of all material wealth; from it we get everything we use or value. All people, at all times, must make use of land. Land parcels are heterogeneous goods, which are usually differing in variety of characteristics and always differentiated at least by location. In other words, land parcels are unique. Land is also a limited quantity, i.e. land exists only in fixed amounts. It cannot be produced like many other products and, as a result, it does not possess manufacturing costs in a traditional sense (developing costs mainly occur when the land use is altered). Theoretically, land has an infinitive duration, i.e. it is theoretically indestructible, which increases its attractiveness as an investment opportunity; land can, however, in certain cases be contaminated that limits its durability. Land markets typically face significant constraints from the economic environment, such as taxes or regulation on land uses.

Land forms an essential part for all human activities and cannot be replaced by any other commodity; land is a necessary factor of production, essential to the provision of urban housing services, the production of agricultural goods and the conduction of business enterprises. In other words, land is required directly or indirectly in the production of all goods and services. At the same time land is demanded as a financial asset, which is often a good hedge against inflation [Brandao & Feder, 1995]. Land is always fixed to a particular location and, therefore, cannot be transferred to another market area, i.e. land is spatially immobile. This implies that location is an intrinsic attribute of land. In reality, there is no such thing as a land market: there are many different ones or, to put it another way, the land market is segmented. In land markets no central trading market or central price listing exists, but lots of local markets typically with low product volumes.

In economic sense, land markets are highly imperfect; it is generally thought that the land market is less efficient than financial markets because of unstandardized commodities, implicit rents, transaction costs, carrying costs and taxes. In essence, land markets are allocatively, informatively and operationally inefficient. Transactions costs play a prominent role in land markets and they are usually quite substantive in monetary terms; transaction costs typically consist of search costs, bargaining costs and enforcement costs. Assets in land markets are generally less liquid than assets in financial markets. Another common feature of the land market is the "lumpiness" of changes in price making forecasting somewhat more hazardous than for markets that respond in a more controlled way to changes in the market area. Moreover, the price formation mechanism in land market tends to be nonlinear and the form of this nonlinearity is usually unknown a priori. In table 3 is contrasted the properties of land markets with that of the perfect capital markets [Virtanen, 2000, p. 14].

Table 3. Dissimilarities of perfect capital markets and land markets.

Perfect capital markets	Land markets
There exist numerous buyers and sellers in the market so that a single buyer or seller does not have an overriding effect on the prevailing market price. Individual buyers and sellers are price takers.	There exist only a few buyers and sellers in the land market. Typically the seller has a monopoly or oligopoly position. Buyers and sellers are price setters.
The product offered in the markets is homogeneous (i.e. of uniform quality) so that each single product is completely replaceable by another product.	Products are heterogeneous or differentiated. Each land parcel is unique and fixed to a particular location. The possibility for perfect substitution between products is very rare or does not exist at all.
Buyers and sellers have a complete information about market prices and offers and a perfect freedom of deciding what products to buy and sell.	Buyers and sellers usually possess an imperfect knowledge about market prices and of their relation to the quality of product and its location.
All participants have a free entrance to the market as a seller or buyer; also they have a freedom of exiting the market.	Entry to the markets is limited. The obstacles may be physical (fixed location) or administrative (e.g. planning).
Market prices are relatively constant and moderate. They tend to gravitate to an equilibium position.	Markets prices fluctuate heavily and tend to increase constantly.

Land prices that vary spatially and over time are needed to determine the optimal allocation of land resources among different uses, as well as to define the value of land as collateral for credit or to define land taxes. Therefore, misperceived land prices could lead to inadequate allocation and use of land resources. Land prices will be capturing all the factors that affect the performance of land markets. It is important to understand that land market prices will not necessarily reflect the value of land; however they typically give a useful approximation to the market value of land. Land prices in the local areas are not uniform nor do they decline with distance from the centre in the orderly fashion that the older theories of land use would predict - the complex contours of the land price maps within local areas reflect the fact that environmental amenities of all kinds also impact on land prices. The interaction of land market demand and supply determines land prices. On the demand side population growth, income and level of economic activity determine how much land is demanded for different purposes. Land supply is determined by topography and physical conditions, patterns of land ownership, availability of infrastructure and government regulations. The total supply of land is fixed (with the exception of reclaiming land from the sea), but the supply of specific land type possesses some elasticity in time. [Trivelli, 1997; DTZ Pieda Consulting, 2002; Serra et al., 2004]

The special characteristics of land markets influences directly to the choice of the research method when land prices are studied. Because land parcels are heterogeneous the direct comparison of different parcels is, strictly speaking, impossible. What we need is a method, which adjusts for quality differences between different land parcels, and thus enables the comparison of different parcels. This is the basic idea underlying the hedonic method, which produces quality-corrected estimates for land prices. The nonlinear nature of land prices and the fact that the exact price formation mechanism is typically unknown a priori makes a strong case for nonparametric and semiparametric modelling methods that are flexible not imposing such a restrictive functional form in advance as the common parametric approach does. In land markets prices do not usually gravitate to an equilibrium position, i.e. land prices are typically continuously evolving, which necessitates that the research method should be able to accurately tract the changes that occur through time.

2.3 Foundations of the Hedonic Approach

The reasoning behind the hedonic approach can be made explicit by a model due to Isakson [2002]. First, let us assume that the subject property is characterized at time t by an attribute vector $\mathbf{x}_t = (x_{t1}\ x_{t2}, \ldots, x_{tk}\ d_{t1}\ d_{t2}, \ldots, d_{ts})'$ which consists of k quantitative attributes x_{tj} and s qualitative attributes d_{tj} and that an $n \times 1$ observation vector of transaction prices of comparable properties, denoted by $\mathbf{p}_t = (p_{1t}, \ldots, p_{nt})'$, is available at time t. Furthermore, if the information on the attributes of comparable properties is collected into the $n \times (k+s)$ data matrix $\mathbf{X}_t$, then the n individual indicated values at time t are:

$$\mathbf{v}_t^{ind} = \mathbf{p}_t + \left(\mathbf{1}_n \mathbf{x}_t' - \mathbf{X}_t\right)\mathbf{w}_t \quad ,$$

where $\mathbf{1}_n$ is a $n \times 1$ column vector of ones and $\mathbf{w}_t$ is a $(k+s) \times 1$ column vector that adjusts for the quality differences between the attributes of a subject property and the comparables. For the derivation of individual indicated values, the valuer has to come up with the values for the adjustment vector $\mathbf{w}_t$ along with the information set that contains relevant property specific and market information. The hedonic approach provides a systematic framework utilising statistical methodology for assessing the necessary adjustment values commonly known as hedonic prices of a property without the need of weighting indicated values individually. For example, in the classical linear regression model [Schulz, 2003, pp. 12-13]:

$$\mathrm{v}_t^{ind} = \frac{1}{n}\mathbf{1}_n'\{\,\mathbf{p}_t \; + \left(\mathbf{1}_n\mathbf{x}_t' \; - \; \mathbf{X}_t\right)\mathbf{w}_t\,\} \;\; ,$$

where v_t^{ind} is scalar and $\mathbf{w}_t = \left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1}\mathbf{X}_t'\,\mathbf{p}_t$.

The fundamental idea underlying the hedonic approach is the hedonic hypothesis: Each property can be characterized by the set of all its utility-bearing attributes or characteristics [Rosen, 1974]. The implication is that an observed price can be modelled statistically in terms of a small number of fundamental characteristics rather than analysed in terms of a large number of similar properties.

Hedonic hypothesis therefore implies the following statistical model for the observed prices [Brachinger, 2003]:

$$p_j = f(\mathbf{x}_j;\boldsymbol{\beta}) + \varepsilon_j \,.$$

Thus an observable price p_j is a random variable. $\mathbf{x}_j = (x_{j1}\, x_{j2}, \ldots, x_{jk}\;\, d_{j1}\; d_{j2}, \ldots, d_{js})'$ is a $(k+s)$-vector of quantitative and qualitative characteristics values. $\boldsymbol{\beta}$ represents an unknown $(k+s)$ parameter vector whose estimates $\hat{\boldsymbol{\beta}}$ are the hedonic or implicit prices of a property, and ε_j denotes the stochastic error term.

In the theory of hedonics the concept of a homogeneous property is an ideal type that does not empirically exist. Properties are differentiated, or heterogeneous, and what we can observe are variants of the idea of a certain property all of which are more or less different [Brachinger, 2003]. As a consequence of the non-existence of a homogeneous property, the price of a property does not exist either. The main idea is that the price $\mathbf{P}$ of a property is a latent variable and the values of this price vary over time. In statistical terms, the price $\mathbf{P}$ of a property is assumed to be a nonstationary discrete stochastic process $\mathbf{P} = (\,p^t\,)_{t\in T}$ as defined by the hedonic regression [Brachinger, 2003]:

$$p^t = f(E(\,\mathbf{x}^t);\boldsymbol{\beta}^t) \; + \; \upsilon^t \,.$$

$\mathbf{x}^t = (x_{j1}^t\, x_{j2}^t, \ldots, x_{jk}^t\;\, d_{j1}^t\; d_{j2}^t, \ldots, d_{js}^t)'$ represents, at time t, $(k+s)$ characteristics random vector that varies over a population of variants and $E(\,\mathbf{x}^t)$ is its expectation [Brachinger, 2003]. Parameter vector $\boldsymbol{\beta}^t = (\beta_0^t\; \beta_1^t, \ldots, \beta_k^t\;\; \gamma_1^t\; \gamma_2^t, \ldots, \gamma_s^t)'$ is a time-specific.

Now an exact definition of the price of a land parcel can be given. For any land parcel, the hedonic price Π^t at time t is the expectation of a stochastic process at time t:

$$\Pi^t \equiv E(p^t) = f(E(\mathbf{x}^t);\boldsymbol{\beta}^t)\,,$$

in which the expectation is taken over all the variants at time t. Hedonic price Π^t of a land parcel is then:

$$\Pi^t(E(\mathbf{x});\boldsymbol{\beta}) \equiv f(E(\mathbf{x});\boldsymbol{\beta}) \; .$$

Hedonic quality of a land parcel at time t is $E(\mathbf{x}^t)$ and hedonic quality of a land parcel is $E(\mathbf{x})$.

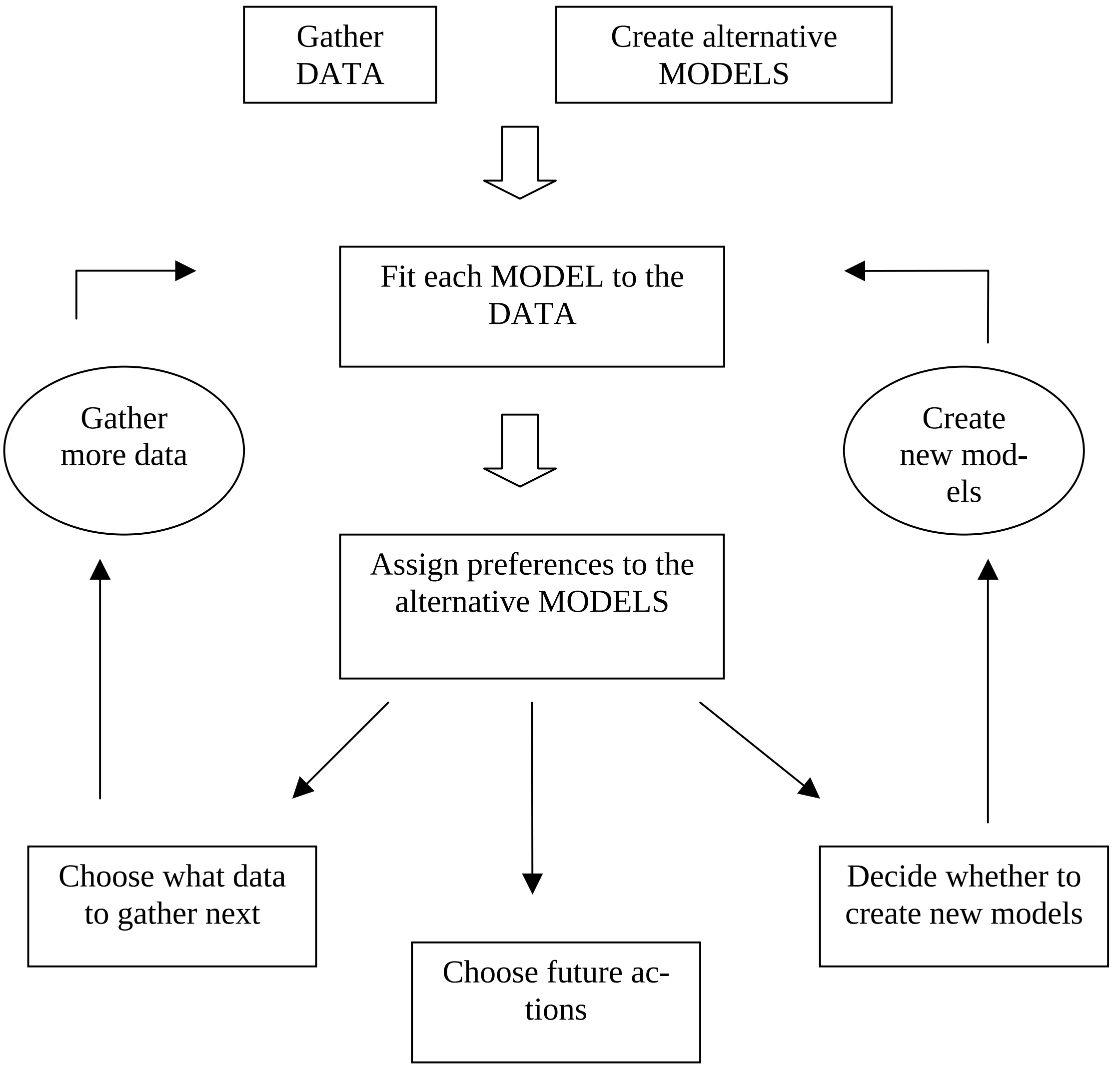

Figure 2. Abstraction of the hedonic modelling process [MacKay, 1992].

In figure 2 is outlined the necessary steps in the hedonic modelling process, where the main philosophy is to evaluate the relative plausibilities of several alternative models in the light of the single data set we actually observe. We start by gathering observations and creating models to account for those data. There are two levels of inference. At the first level "fitting each model to the data", the task is to infer what the hedonic prices of each model might be given the data. The second level of inference is the task of model comparison. After the hedonic models have been fitted and compared, we can then decide to gather more data or to invent new models for the data, and we can repeat the inference process. We might also use the knowledge we have gained from the observations to make decisions about our future actions. [MacKay, 1992]

The starting point in a hedonic analysis is the previous empirical experience and theoretical knowledge concerning the phenomenon that should assist in the formulation of a specific hedonic model. However, the accumulated body of previous empirical experience always relates to specific submarkets, time periods and land types which usually differ from the one we are analysing, and thus this information cannot give a definitive answer to model formulation problem. Also theoretical knowledge is typically insufficient and only gives some broad guidelines how to specify a hedonic model. Therefore the model specification heavily rests on the performance of a model (e.g. how well it predicts the new cases) on the data set we actually possess. In a parametric modelling scheme it is desirable that we work on a set of different model specifications and try to find the one that matches the evidence in most measurable respects. Nonparametric and semiparametric approaches allow a greatly reduced attention to a model specification problem, because they can automatically learn much of the relevant model structure from the available data set with a minimal set of a priori assumptions.

During the 1950's, 1960's and 1970's the emphasis in econometrics was on efficient estimation of a given hedonic model. Nowadays the emphasis has shifted to specification testing, diagnostic checking, model comparison, and an adequate formulation of the expectational variables given the pervasive role of expectations in almost all economic theories. Today every time we estimate a hedonic model (or preferably a set of hedonic different models), we typically test whether it has a significant specification error. If a significant specification error exists, we have to invent and estimate a new hedonic model. However, if the hedonic model is adequate then we can use it to explain the formation of land prices and to predict land prices.

Traditionally (in the parametric modelling context) hedonic analysis can also be classified into the four phases [Hiltunen, 2003, p. 37]:

1. Construction of a hypothesis and the formulation of a hedonic model.
2. Estimation of model parameters with a suitable estimation technique.
3. Investigation of the reliability of the estimated parameters.
4. Study of the predictive ability of the estimated model.

The hedonic analysis process is also very similar in the nonparametric and semiparametric modelling context. However, in many cases with nonparametric approaches the 3. phase is omitted as the parameters usually have no intuitive interpretation.

2.4 Cointegration and Error Correction Models

In the research papers 1 and 3 the concept of cointegration and equilibrium correction models are important. One of the main virtues of equilibrium correction models is that they can effectively separate between short-run and long-term dynamics. It has been empirically observed that equilibrium correction mechanisms can increase the explanatory power of a given hedonic model and usually uncover new relationships present in the data [Tulaca et al., 2000]. There is an established isomorphism between equilibrium correction models and cointegration, i.e. one implies the other [Engle and Granger, 1987]. In essence, if there exists a valid equilibrium correction mechanism for the variables in the series they have a long-run equilibrium relationship, or steady state, to which the time series eventually adjusts.

A time series is said to be cointegrated if the series is nonstationary with a unit root but some linear combination of the series is stationary [Chaudhry et al. 1999]. Suppose that we have a simple regression equation $y_t = \beta x_t + \varepsilon_t$ and that $y_t \sim$ I(1) and $x_t \sim$ I(1) where I(.) denotes a degree of integration. Then y_t and x_t are said to be cointegrated if there exists a β such that $y_t - \beta x_t$ is I(0). What this means is that the regression equation $y_t = \beta x_t + \varepsilon_t$ makes sense because y_t and x_t do not drift too far apart from each other over time. Therefore there is a long-run relationship between them. By asking the question of whether y_t and x_t cointegrated, we are asking whether there is a long-run relationship between the trends in y_t and x_t. The fundamental aim of cointegration analysis is to detect any common stochastic trends in the data, and to use these common trends for a dynamic analysis of a series.

2.5 Data of the Study

Empirical studies of land prices tend to exhibit significant sensitivity to changes in data, i.e. different submarket, time period and land use can lead to widely differing results, even in the context of unified methodology. The majority of that variation is explained by spatio-temporal movements: functional forms and parameters tend to vary with location and are not homogeneous throughout the data set, whereas temporally changing market conditions cause data-generating processes to evolve over time. To reduce the sample dependency, i.e. to improve the invariance of empirical studies, this dissertation examines two data sets[3] that are located in different submarkets and associated with partially non-overlapping time frames. The land type (land use) is, in contrast, fixed in order to reduce unnecessary heterogeneity of land prices. It represents undeveloped land that has not yet reached its highest and best use: vacant sites without a local detailed plan that are reserved for residential housing purposes.

The first sample data involve observations on land prices and the associated characteristics in the municipality of Espoo, a highly polycentric city, which lies inside the Helsinki metropolitan area with circa 225 000 habitants; its population is the second largest of the cities in Finland, which has experienced a rapid growth in its late history. The study period is from January, 1990 to December, 2001 with a total number of observations of 400[4]. The observations from the last year (total of 39) are held back for post-sample predictive testing. In terms of quality, this data set is preferable over the second sample, i.e. it has been pre-checked for any errors and analysed before. In table 4 are documented some standard sample statistics for the study variables in the case of Espoo.

The second sample contains observations of land prices and the associated attributes on 793 unbuilt land parcels without a local detailed plan sold in the period spanning from January, 1985 to March, 2004 in the municipality of Nurmijärvi, which lies just outside the Helsinki metropolitan area with approximately 36 000 habitants and three distinctive population centres (the parish village, Klaukkala and Rajamäki). Nurmijärvi has recently also witnessed several years of rapid expansion. The last 50 observations were left aside for evaluating the predictive accuracy of estimated local models outside the estimation sample, which corresponds ap-

[3] In the paper 1 only one data set (the Espoo case) is investigated, whereas in the papers 2-5 two data sets (the cases of Espoo and Nurmijärvi) are analysed.
[4] The paper 1 uses, in fact, 405 observations from the submarket of Espoo. In the papers 2-5 400 observations from the submarket of Espoo are used (in these papers 5 observations were dropped as clear and strong outliers from the subsequent analysis). Also in the Nurmijärvi case the original number of observation was over 800, but some dubious observations were dropped from the final analysis.

proximately to a one-year period from the late February 2003 to the early March 2004. In terms of quantity, this data set offers more opportunities to flexible and partially flexible modelling than the Espoo case, albeit it is more pronounced to any errors (e.g. recording errors), since it has not been pre-checked for hedonic modelling purposes. In table 5 are documented some standard sample statistics for the study variables in the Nurmijärvi case.

Table 4. Some common sample statistics for the municipality of Espoo.

Variable (unit)	**Arithmetic mean**	**Minimum**	**Maximum**	**Std. Deviation**
Total price (€)	59126.40	3027.00	756846.00	61976.88
Square price (€/m^2) (unit price)	22.99	0.24	127.55	19.09
Parcel size (m^2)	4207.49	1000.00	28400.00	4613.75
Distance to CBD of Helsinki, L_2-metric (km)	17.22	7.61	27.29	4.34
Northing	667922.61	666540.00	669650.00	695.24
Easting	253796.25	252815.00	254638.00	406.98
Quarterly price index of single-family houses	154.06	116.80	187.30	22.35
Parcel type (=0 if whole site; 1 otherwise)	-	0	1	-

Table 5. Some common sample statistics for the municipality of Nurmijärvi.

Variable (unit)	**Arithmetic mean**	**Minimum**	**Maximum**	**Std. Deviation**
Total price (€)	22 019.03	673.00	479 336.00	21 262.25
Square price (€/m^2) (unit price)	3.80	0.34	22.83	3.27
Parcel size (m^2)	7387.36	1 000.00	30 000.00	4 651.66
Distance to CBD of Helsinki (km)	33.11	22.28	44.91	5.22
Distance to parish village of Nurmijärvi (km)	8.55	0.32	16.06	3.37
Distance to population center of Klaukkala (km)	9.33	0.50	21.44	5.00
Distance to population center of Rajamäki (km)	11.86	0.43	20.92	4.58
Northing	670354.00	669240.00	671736.00	600600.00
Easting	254091.50	253031.00	255695.00	250546.70
Quarterly price index of single-family houses	154.30	100.00	226.00	32.00
Parcel type (=0 if whole site; 1 otherwise)	-	0	1	-

It should be noted that during the last 20 years the Finnish economy has witnessed some dramatic changes that have possessed an influence on land prices also. The Finnish money markets were liberated in the middle of the 1980's, which resulted to a strong economic growth and, which reduced the number of unemployed. This also led to the overheating of the whole Finnish economy. The Finnish economy witnessed a strong depression in the beginning of the

1990's. The real gross national product dimished 10-12 % in the years 1991-1993. At the same time the unemployment rate rose to almost 20 %. Our economy started to recover in 1994, after which the real gross national product was for seven years well over the average value in the EU-countries. It should be noted that changes in the state of economy transfers first to the housing markets and after that to the site markets. These unique events makes it hard for the results of this study to generalise outside the sample period and thus the results should be understood mainly in the context of the investigated time interval (and in the context of analysed submarkets and land type). Different time periods - not to mention different submarkets and land type - tend to generate, at least partly, different results and conclusions.

3 Modelling of Temporal Variation

3.1 Concept of Structural Time Series Models

The variation of observed land prices is a combination of cross-sectional and time series variations [Schulz, 2003, p. 58]. An empirical model of land prices has to recognize these two different, yet closely related sources of variations. Besides the spatial characteristics, the selling date is an important attribute in explaining the evolution of market prices through the flux of time which itself is directly an unobservable quantity, i.e. time is a latent variable. What we can observe are different states that occur in a predefined submarket and changes that they cause in prices in that market area. Temporal variation is a result of changing market conditions, which are driven by, among others, changes in consumers' preferences, investors' expectations, technological advantages, income changes and interest rate changes. The temporal variation can be understood as representing that part of price variation that is more or less common to all parcels of land in the same submarket.

In modelling the time series or temporal variation of land prices it is important to understand that the behaviour of land prices over time, which is also typical of wider range of economic time series, is generally nonstationary or transient meaning that the data-generating process itself evolves over time. More specifically, nonstationarity denotes the general sense of processes whose first two moments (conditional expectation and the variance of its error distribution) are not constant over time[5]. This dynamic nature of data-generating processes is attributable to changes in economic environments, technological progress, political shifts, cultural movements, etc.

The effect of temporal variation is also multidimensional: Often one can legitimately separate the price trend, the price cycle, the seasonal price variability and the irregular price variability from each other. The trend can be understood as that part of the series when extrapolated gives the clearest indication of the future long-term movements; it can be linear or nonlinear. The simplest choice of a trend would be a deterministic linear time trend, but this usually is too restrictive, unless the time period is very short [Harvey, 1997]. Cycles are characteristic to many economic time series as economy goes from boom to recession and back again. More specifically, the cycle refers to the ups and downs seen somewhat simultaneously in most parts of a

[5] There are in fact two common definitions of stationarity. Weak or covariance stationarity refers to the situation where the first two moments (mean and variance) of the series are time-independent, whereas strict stationarity refers to the situation where all moments (not just the mean and variance) are constant. In this study stationarity refers to the weak stationarity and thus nonstationarity is the situation where the first two moments of the series are not constant in time.

land market; it involves shifts over time between periods of relatively rapid growth of land prices alternating with periods of relative decline. Seasonals represent patterns of change in a time series within a year; they tend to repeat themselves each year. Irregular price variability is the unexplainable or random variability of the price series.

The nonstationary behaviour of land prices and the multidimensional nature of time cause that the valid hedonic modelling of land prices is in practice a very challenging task. Traditionally in the land value studies [see e.g. Hiltunen, 2003] the modeller has tried to reduce the temporal variability of land prices to the variability of consumer price index or to some relevant house price index measure, which have subsequently been used as explanatory variables in the hedonic regression. The main problem of this approach is that the effect of time will usually be represented inadequately for practice purposes since the indexes are aggregate measures that can trace temporal changes only in very limited sense; therefore, much of the relevant time variability will be left unexplained and thus ignored. Also the temporal variation of land prices has traditionally been often explained by economical time series such as income levels, interest rates and population changes. These measures share the same criticism as the consumer price and house price index measures. Trend removal by differentiating the price series and using ARIMA (autoregressive integrated moving average) models has been proposed many as a general framework to handle temporal variability in economic modelling. The process of differentiating the series, unfortunately, leads to a serious loss of potentially valuable long-term price information. Furthermore, the ARIMA scheme inherently involves the assumption that stochastic processes are stationary or, more specifically, can be transformed to stationary series by the process of differencing them (possible multiple times).

Perhaps the most popular solution in modelling the temporal dimension of land prices has been the indicator variable technique, i.e. to use time dummy variables to account for the effects of time. Albeit widely applied, this approach has a multitude of serious drawbacks. The core problem of using the indicator variable procedure seems to be its inflexibility as the effect of time is merely represented by a series of fixed discrete jumps, which tend to be inaccurate in practice. The second problem is the lack of sufficient degrees of freedom, since estimation involves an extensive set of time-indexed dummy variables along with other regressors, at least one for each time period. The third problem is the necessity to choose the correct time interval, e.g. a period of one-year, which somehow reflects a typical decision-making horizon, although economic agents operate and actions take place simultaneously at various different time scales. The final major limitation of dummy variable technique is that the estimated model structure potentially suffers from the acute multicollineary problem, which distorts the estimated parameters, when used to represent the time evolution of land prices.

There are several other and more sophisticated approaches to account for the effects of time variability than e.g. the indicator variables technique or ARIMA models in the hedonic context. The structural time series approach is often a viable tool, which can separate long-term price movements (trends and cycles) from seasonal and irregular price variability. They are suitable for the analysis of nonstationary features of price series, in which time interval need not be equispaced (i.e. a time series is simply a set of observations ordered in time). In essence, structural time series models can be thought of as a certain type of generalized regression models in which explanatory variables are functions of time and the parameters are time-varying [Harvey 1989, p. 10; Harvey & Shephard, 1993; Harvey, 1997]. More precisely, structural time series or unobserved component models can be understood as semiparametric estimators that combine many of the benefits of parametric and nonparametric estimators; temporal variability of land prices is estimated in a nonparametric fashion, which permits the effect of time to be linear, convex and concave in different regions, whereas the hedonic prices of attribute variables are estimated in a parametric manner. It is thus expected that the parameters for exogenous variables can be estimated more accurately using these models. In a structural model an explicit stochastic trend is assumed in which the level and slope coefficient are allowed to evolve over time. When using unobserved components cycles are modelled effectively by means of a mixture of sine and cosine waves.

When considering the determination of hedonic prices in land markets and, specifically the temporal dimension, there are several benefits in using the structural time series approach and the associated state space form as compared to the Box-Jenkins ARIMA methodology. These include [Harvey & Shephard, 1993; Harvey, 1997; Durbin & Koopman, 2002, p. 51-53]:

- *Structural analysis of the problem*. Different components that make up the series, including the regression elements, are modelled explicitly when, in contrast, the Box-Jenkins approach is a sort of "black box". A structural model provides not only the forecasts of the series but also presents a set of stylised facts. Also a structural model can be handled within a unified statistical framework that produces optimal estimates with well-defined properties.
- *Management of nonstationarity*. In a structural model nonstationarity (transitory parts of the model specification) can be handled conveniently by unobserved components without the need of differencing any variables. By comparison, in the Box-Jenkins approach the stationary is assumed, and nonstationary components of the series are usually eliminated by differencing the variables, which results to a potential loss of valuable long-term information. Furthermore, the standard unobserved component models are simple, yet effective, leading to parsimonious representations for the systems.
- *Generality*. Multivariate observations can easily be handled with structural models, which cover as special cases a wide range of econometric models (including all ARIMA models). Explanatory variables can be introduced into the model structure and the associated regression coefficients (hedonic prices) can be permitted to vary stochastically over time if needed. Different kinds of intervention variables, e.g. impulse and level interventions, can be specified and lagged values of dependent as well as explanatory variables can be incorporated to a model. Missing observations and varying dimensionality of observations are issues that are straightforward to deal with structural models.

3.2 Estimation Results of Structural Time series Modelling

Structural time series models have been investigated in two single localities of the Finnish land market (submarkets of Espoo and Nurmijärvi) in the paper 2 "An Analysis of Land Prices: A Structural Time-Series Approach" [Hannonen, 2005b]. To improve theory-dependency of this analysis, the empirical section considers estimation of two different common econometric models of land prices - (i) the classical linear model and (ii) the multiplicative form of double-log model - in the context of structural time series modelling formalism. To reduce sample dependency, the paper examines two distinct data sets that are located in different submarkets and associated with partially non-overlapping time frames. The land type (land use) is, in contrast, fixed; it represents undeveloped land not yet reached its highest and best use, i.e. vacant sites without a local detailed plan that are reserved for residential housing purposes.

Unobserved component models of the study are estimated using the Kalman filtering and smoothing techniques. The importance of the Kalman filter is based [Koopman et al., 1999]:

- Computation of one-step ahead predictions for observations and state vectors with the associated mean square errors;
- Diagnostics via one-step ahead prediction errors;
- Computation of likelihood function by means of prediction error decomposition;
- Smoothing that uses the outcome of the Kalman filter.

The importance of the Kalman smoother is based [Koopman et al., 1999]:

- Signal extraction, detrending and seasonal adjustment;
- Diagnostics for separating the pieces of information between outlying observations and structural changes via auxiliary residuals;
- EM-algorithm for initial estimation of parameters;
- Calculation of the score vector.

One of the key attractions of structural time series modelling formalism is that it combines the flexibility of a time series model with that of the interpretation of a regression analysis. Structural time series models of the study consisted of a description of regression effects, trend component, cycle component and first-order autoregressive process for disturbances. Regression effects were measured by two common hedonic models, a linear model and a double-log model, in order to reduce the regression specification error. Trend component was estimated by two different trend models, a local level model and a local linear trend model, in order to reduce the trend specification error. Cycle component consisted of 1 to 3 different cycle terms and was estimated by a mixture of sine and cosine waves. First-order autoregressive process for disturbances can be seen as limiting case of a stochastic cycle.

The empirical investigation [Hannonen, 2005b] evidenced that quite plausible trend and cycle specifications could be identified in most model formulations for regression effects (3 out of 4 different formulations for regression effects), which resulted significant improvements in their post-sample predictive validity. The one exception was the case of classical linear model structure for regression effects in the municipality of Nurmijärvi, in which the conventional modelling approach outperformed the structural one. Overall, these improvements in the post-sample predictive accuracy were more pronounced and coherent in the municipality of Espoo, whose data were screened and pre-checked for any errors prior to the analysis.

The explanatory power of different model contenders in Espoo and in Nurmijärvi cases ranged from 0.61 to 0.70 in the case of Espoo and was 0.46 in the case of Nurmijärvi when the fit was measured by the standard coefficient of determination and ranged from 0.77 to 0.83 in the case of Espoo and from 0.68 to 0.70 in the case of Nurmijärvi when the fit was evaluated by the modified coefficient of determination. Goodness-of-fit statistics indicated that the conventional double-log model for regression effects yielded to a more data-congruent specification than the classical linear model[6]. In that model structure, beyond any reasonable doubt, the two most important determinants of land prices were the estimated trend term and parcel size vari-

[6] In the Nurmijärvi case the coefficient of determination statistics that relate to the double-log model and the linear model are very similar. However, AIC and BIC statistics indicate that double-log model is a slightly better model choice for regression effects in the Nurmijärvi case. In the Espoo case all goodness-of-fit statistics indicate that the double-log model is more data-congruent specification for regression effects than the linear model.

able. The house price index variable confounded the analysis of unobserved component models, when transformed (logarithmic) land prices were analysed.

The simplest trend model, the local level model, seemed to be the most appropriate description of the trend component in all model formulations. In a more complex trend model, the local linear trend model, the slope coefficient was statistically insignificant and therefore added no new information when compared to the local level model. The behaviour of estimated trends was more plausible in the Espoo case, where the price trend component oscillated approximately in the range of (12 €, 18 €) in the study period. In the submarket of Nurmijärvi, the estimated trend model tended to overshoot the influence of trend component, which fluctuated around (4.5 €, 11 €), which is significantly higher than the average value of the series in that period.

One to three different cycle components were identified in all model formulations. These all seemed to capture, more or less, a high-frequency component of land price movements, and as such, may not be very informative. However, some resonance was discernible in some cycle models. The effects of estimated trends and cycles were in most cases highly curvelinear and quite volatile. The first-order autoregressive process for irregular disturbances was found to be an integral part of the overall model in three (out of four) different cases. The occurrence of stochastic trends and/or cycles in the analysed cases suggests that these markets are not in the steady state.

The data analysed contained many outlying observations in terms of an unusual high value of standardised residual, which distorted the analysis - especially in the case of Nurmijärvi. Instead of removing the outlier its effect was statistically measured by an impulse intervention variable and the influence was subsequently included as part of the overall model specification resulting to no loss of price information.

When using the conventional double-log specification for regression effects and the local level model for trend specification in the Espoo case, the pronounced dip in the series corresponds to the first quarter of the year 1995 after which the series has grown upward for circa five years. The dominant peak in the Espoo case occurs at the beginning of the series: in the second half of the year 1990. When the usual double-log specification for regression effects and the local level model for trend specification are used in the Nurmijärvi case, the dominant peak in the price series corresponds to second half of the year 1990. The dominant dip takes place about the turn of the years 1995 and 1996 after which the series has moved upward about seven years reaching the high price level of the year 1990. The estimated hedonic prices and unobserved components for different model formulations and submarkets are represented in table 6. In essence, there is a significant variability between different hedonic models and submarkets.

The implication of this investigation with unobserved components is that the structural time-series modelling paradigm offers a viable alternative to the conventional hedonic approach based on the ordinary least squares. In the conventional approach the accurate measurement of time element is usually problematic and no consensus exists how to handle the time component. In the structural approach the multidimensional effect of time can be measured and handled conveniently by different unobserved components that provides estimates for a trend term and a cycle term. The empirical investigation showed that in most model formulations the use of unobserved components significantly improved the post-sample predictive accuracy when compared to the orthodox approach.

Table 6. Estimated hedonic prices and unobserved components.

	Submarket of Espoo		Submarket of Nurmijärvi	
Variable	Linear model	Double-log model	Linear model	Double-log model
Level	18.28 [3.62]	12.64 [27.97]	5.18 [3.26]	10.78 [18.87]
AR(1)	-2.46 [NA]	-1.77 [NA]	-1.96 [NA]	-
Cycle 1 (comp. #1)	-1.88 [NA]	0.032 [NA]	-0.31 [NA]	0.014 [NA]
Cycle 1 (comp. #2)	0.86 [NA]	0.084 [NA]	0.13 [NA]	0.026 [NA]
Cycle 2 (comp. # 1)	-4.78 [NA]	-0.20 [NA]	-	-0.0051 [NA]
Cycle 2 (comp. #2)	-0.97 [NA]	-0.069 [NA]	-	0.018 [NA]
Cycle 3 (comp. # 1)	-0.041 [NA]	-	-	-
Cycle 3 (comp. #2)	1.63 [NA]	-	-	-
Parcel size	-0.0014 [-10.22]	-0.76 [-1960]	-0.00028 [-14.97]	-0.62 [-1892]
Distance 1	-1.90 [-12.41]	-1.31 [-9.59]	-0.12 [-7.53]	-0.93 [-7.29]
Distance 2	-	-	-0.096 [-3.83]	-0.22 [-5.12]
House price index	0.28 [8.58]	-	0.038 [5.30]	-
Parcel type	-	-0.31 [-5.15]	-0.75 [-4.18]	-0.21 [-5.03]
Intervention #1	68.93 [6.27]	-2.70 [-4.95]	13.37 [5.88]	-2.34 [-4.28]
Intervention #2	51.34 [4.67]	-	13.74 [6.03]	-2.81 [-5.11]
Intervention #3	46.40 [4.23]	-	11.68 [5.13]	-
Intervention #4	-	-	13.03 [5.73]	-
Intervention #5	-	-	11.84 [5.20]	-
Standard coefficient of determination	0.61	0.70	0.46	0.46
Modified coefficient of determination	0.77	0.83	0.70	0.68
Number of observations	400		793	

t-values are presented in the brackets.

In the structural time series approach, the observations are directly made up of interpretable components of trend, cycle, seasonal, and regression term plus error. It therefore provides a wide class of information on different hedonic modelling aspects, not merely the regression effects. When considering the effect of time, the use of unobserved component models can improve our understanding of the nature of long-term land price movements and short-run land price fluctuations. The use of unobserved components is not, however, without problems.

The data sets analysed in this study are heterogeneous and many different outlying and influential observations existed in the data. These tended to distort the estimated trends and cycles even when impulse intervention variables were used; In particular the estimated trend was too high in the Nurmijärvi case. Furthermore, the time period that was analysed in the Espoo and Nurmijärvi case, contained highly exceptional market conditions (overheating of the economy and/or a deep depression) that makes it difficult to generalise the results of study to different times.

3.3 Concept of Wavelets

Structural time series models seem to posses a serious disadvantage, if the observable series contain many outlying or influential observations as is typical of many land value studies; the unobserved component estimates tend to overshoot (undershoot) the effect of levels[7]. In those cases, the estimated trends and cycles can be non-informative per se, i.e. they fluctuate at too high or low a level to be meaningful as such, although they are an important part of the overall model structure in conjuction with the estimated hedonic prices. Fourier-based methods are also one interesting flexible alternative to the standard indicator variable method or to the ARIMA approach. However, they have a hard time in reproducing the nonstationary elements of the price series that are common in land markets. In essence, wavelets that are non-parametric orthogonal series estimators extend the main ideas of Fourier analysis to situations, in which different forms of nonstationary behaviour (e.g. trends, abrupt changes or chirps and volatility clustering) are expected. The paper 4 "An Analysis of Trends and Cycles of Land Prices Using Wavelet Transforms" [Hannonen, 2006b] considers the applicability of wavelets in the land markets.

The application of wavelets has rapidly increased over the last 20 years with over 1000 reviewed papers now appearing each year, and with the total number of over 16 000 articles being published to date in diverse fields of study [Addison, 2004; Crowley, 2005]. The theoretical underpinnings of wavelets were completed in the late 1980's, whereas the 1990's witnessed a rapid increase in the number of different practical applications. These applied fields include [Graps, 1995], among others, signal and image processing, data compression, astronomy, acoustics, fractals, partial differential equations, medicine, seismology, speech discrimination, optics and nuclear physics. At the moment they are at the verge of entering the mainstream econometrics [Schleicher, 2002; Gencay et al., 2002] with some applications in different fields of finance and economics [Ramsey, 2002]. In land markets wavelets represent a totally new perspective of time series modelling, whose empirical performance has not been investigated so far.

There are at least four different uses of wavelet analysis in the land market, which are [Ramsey, 2000; Crowley, 2005]:

(1) Explanatory analysis to gain new insight of the data or phenomenon;

(2) Density estimation and regression for evaluating spatially inhomogeneous surfaces;

(3) Time-scale decomposition of disaggregate series;

[7] The effect of unusual observations can, however, be mitigated by the proper use of different intervention (e.g. impulse intervention) variables.

(4) Disaggregate forecasting.

In explanatory analysis the relevant question that can be tackled with wavelet estimators is the time scale versus frequency: in land economics an examination of data to evaluate the presence and ebb and flow of frequency components is potentially valuable. In density estimation and regression wavelet estimators are superior to conventional kernel estimators, whenever local inhomogeneities are present (which are highly typical in the land market). In time-scale decomposition the crucial issue is the recognition that meaningful relationships between land economic variables can possible be found at a disaggregate (scale) level rather than an aggregate level. In disaggregate forecasting wavelet estimators provide a basis for establishing global versus local aspects of the price series, therefore addressing the question of whether the forecasting is really possible. All of these can potentially lead to different kinds of gains, including improvements in the bias-variance trade-off, new insightful perspectives to the nature of data-generating process and enchainment of robustness to modelling errors [Ramsey, 2002]. The paper 4 concentrates on the first and third areas of uses in two single localities (Espoo and Nurmijärvi) of the Finnish land market. In order to reduce sample dependency, the paper examines two distinct data sets of undeveloped sites that are located in different submarkets and associated with partially non-overlapping time frames.

The classical Fourier transform's utility lies in its ability to analyse a function in the time domain for its frequency content [Graps, 1995], which has been successfully applied in a healthy amount of different practical applications and theoretical considerations. The main problem of the Fourier approach is, however, that it gives information only about how much of each frequency exists in the function, but it does not tell modeller when in time domain those frequency components appear; in other words, it has only frequency resolution and no time resolution. In practice, the Fourier transform has trouble reproducing transient signals and functions with abrupt changes, such as land prices. This means that the Fourier transform is only applicable to the analysis of stationary processes, whose frequency content does not change in time. In the case of stationary series, one does not need to know at what times frequency components appear, since all frequency elements exist at all times. When the time localization of the spectral components is desired, a transform giving the simultaneous time-frequency representation of the function is needed. [Polikar, 2001]

Wavelet transform, on the other hand, is capable of providing the time and frequency information simultaneously; wavelets are localized both in the time and the frequency domain. This enables them to escape Heisenberg's curse, i.e. the physical principle that tells that one cannot be simultaneously arbitrarily precise with respect to the exact frequency and the exact time occurrence of that frequency of a function [Vidakovic, 1999, p. 35; Schleicher, 2002]. Wavelets are designed to give good time resolution and poor frequency resolution at high frequentcies and good frequency resolution and poor time resolution at low frequencies [Polikar, 2001]. This approach is particularly useful if the function has high frequency bursts for short durations and low frequency components that last a longer period of time. Furthermore, wavelets tend to be much less sensitive to any errors in the data, since they can effectively separate the long-term movements from high-frequency details, whereas in the Fourier transform these errors - that are common in land price studies - can transform a smooth function into a jumpy one and vice versa, which is highly undesirable [Mackenzie, 2001].

There exist several different criteria of how to choose an appropriate class of wavelet functions to represent data. Symmetry of the wavelet basis function is useful for describing signals, which exhibit local symmetries. Orthogonality is a highly desirable property of any transform as it significantly simplifies calculations associated with the transform in question. Smoothness of the wavelet is yet another important characteristic that is measured by the number of continuous derivatives of the basis function. In the empirical section orthogonal, compactly supported (with finite energy) and reasonably symmetric wavelet functions called the Symlets, with eight vanishing moments, were chosen after some empirical experimentation as the proper class of wavelet basis functions to represent the data. The scale parameter is the crucial element in wavelet analysis, which is similar to scale used in maps [Polikar, 2001]. High scales represent low-frequency variation, which give global information about the function, whereas low scales correspond to high-frequency component of variability, which brings out detailed micro-information about the signal. This varying scale enables the researcher, as Graps [1995] states, to see both the forest and the trees. Change of the scale from high to low means, in this context, zooming in and seeing the trees in the structure.

3.4 Estimation Results of Wavelets

The empirical investigation [Hannonen, 2006b] witnessed that the wavelet estimators, which are capable of providing the necessary time and frequency information of land prices simultaneously in a highly flexible fashion, yielded, in all cases, to meaningful and plausible representations for the low-frequency unit land price fluctuations. In the empirical study both the raw or untransformed and quality-adjusted (adjusted by using local regression modelling tools) or transformed unit land price series were modelled using the one-dimensional discrete inverse wavelet transform to account for cyclical patterns and trends that were hidden in the original time-amplitude representation of the unit prices.

The trend estimation of unadjusted unit price series in the Espoo case revealed that the unit land price has fallen approximately 52 %, and quite steadily and almost linearly, in the period spanning from the beginning of year 1990 to the turn of the years 1995 and 1996. The five-year time period of 1996 to 2001 evidence a steady increase in the price level with the rise of circa 63%. The trend estimation of quality-adjusted unit price series in the Espoo case evidenced that there are one dominant peak that occurs in the turn of the years 1992 and 1993 and one trough that occurs at the first quarter of 1999 in the estimated price trend. The trend of unit prices has decreased (28% overall) almost linearly about six years in the period of 1993-1999. In the period of 1990-1992 the price level has increased by about 8% and in period of 1999-2001 by about 17%.

The trend estimation of unadjusted unit price series in the Nurmijärvi case indicated that the reconstructed signal shows a clearly discernible dominant peak and trough in the middle of the series. This dominant peak corresponds quite closely to the turn of the years 1989 and 1990 after which the level of unit prices has fallen systematically about eight years by circa 44%. The dominant valley thus occurs in the turn of the years 1997 and 1998 after which unit price level fluctuations has steadily increased over six years by an impressive 160 %. This should be contrasted with the five-year period of 1985-1989 where the increase in the level of the price series has been about 125%. The trend estimation of quality-adjusted unit price series in the Nurmijärvi case evidenced that there exist a clearly discernible dominant peak and valley in the estimated series; the dominant peak relates to the turn of the years 1988 and 1989 after which the quality-adjusted unit price has fallen steadily over nine years by circa 17%. Since the beginning of the 1998, the quality-adjusted unit price has increased steadily and linearly almost six years reaching the high unit price level of the turn of years 1988 and 1989.

The estimated cycles and trends were all nonlinear and, in particular, the behaviour of cyclical component was highly curvelinear and transient in time. These findings strongly suggest that the submarkets in question are not in a steady state, but are continuously evolving in time. The modelling of the raw and the quality-adjusted unit land price series resulted to different kinds of descriptions for the high-scale temporal variability. It seems that much of the temporal variation present in the untransformed series is, in fact, explained by the quality differences in the attribute variables. As a result the use of unadjusted price series tended to exaggerate the salient features of price fluctuations. The use of quality corrections produced seven to fifteen times smaller internal variability when compared to the original, untransformed series; and also significantly reduced internal variability when contrasted to the unadjusted wavelet-based series. In contrast to conventional methods (e.g. the time dummy variable technique or Fourier analysis) the wavelets estimators can, in principle, handle all sorts of nonstationary price behaviour in the land market, i.e. it can well manage complex, non-periodic, noisy, intermittent and transient signals, whose valid measurement would otherwise prove to be highly difficult.

The main use of wavelet estimators in land markets is in estimating the price trends and cycles. As a wavelet estimator does not perform regression analysis, the price series should be quality-adjusted by some other method. In this study flexible local regression analysis was used to take account of differences in the attribute variables. The crucial element in wavelet analysis is scale; by changing the scale we can see different features of the series. The scale is also a source of problems in wavelet analysis; in practise it is difficult to decide which scale is the correct one. The final choice of scale is typically very ad hoc. Furthermore the choice of the wavelet basis function has to be made; the decision to choose a particular basis function is, to some extent, subjective. Wavelet estimators' main benefit over structural time series models is that they are more flexible and not sensitive to outlying data points.

The time periods analysed in the paper 4 are unusual in that there was a overheating of the whole Finnish economy and/or a subsequent deep depression in those time frames. This makes it difficult to generalise the empirical findings of the paper 4 across time and, as a result, the empirical results should mainly be understood within examined time frames.

4 Spatial Variation

The cross-sectional or spatial variation stems from the differentiated nature of a land depending on certain characteristics such as location, lot size, intended use and planning status of a land parcel. This cross-sectional variation is usually divided into the concepts of spatial heterogeneity and spatial dependence. Spatial heterogeneity implies that functional forms and parameters vary with location and are not homogeneous throughout the data set, whereas spatial dependence implies that the variation is a function of distance. The problem due to spatial dependence can often be resolved by introducing into the hedonic model structure terms that describe location or distance, which are then used as explanatory variables in the given model. Instead, the problem due to spatial heterogeneity is usually more complicated. One natural solution would be to narrow the analyses into very small submarkets, which homogenises the data. However, in practise this operation is not typically feasible due to the scarcity of observations for the hedonic modelling purposes. Adaptive modelling techniques, such as the local regression [see below], usually provide a better solution to the spatial heterogeneity problem in that they possess a spatial adaptation property and thus explicitly address the spatial heterogeneity problem.

4.1 Hedonic Modelling Approaches and the Model Specification Dilemma

Hedonic modelling approaches can be classified by their flexibility in discovering conditional structure to three categories [see, e.g. Pace, 1995]:

(1) parametric approaches;
(2) semiparametric approaches;
(3) nonparametric approaches.

Parametric models that represent data modelling culture [Breiman, 2001] have formed the conventional dogma of hedonic pricing methods in land value studies, where prespecified global models are estimated by means of ordinarily least squares or some modification thereof. Benefits of parametric approaches undeniably include: simplicity, interpretability, parsimony and comprehensive statistical theory. The fundamental obstacle, however, underlying the general use of parametric models is their inflexibility, i.e. inability to learn genuine structure about the hedonic relationship from the evidence in such decision-making settings, where theoretically unknown nonlinearity or nonstationarity is expected. This is the typical case when the effects of variables representing location and time are considered [McMillen and Thorsnes, 2003]. The conventional result is that even the best parametric model tends to impose restrictions that substantially reduce the explanatory and predictive power of hedonic equation [Pace, 1993 & 1995; Anglin & Gencay, 1996; inter alia]. Unless the theory-laden parametric model coincides with the data-generating process (this is a strong assumption), profound misspecification errors may result imposing serious threats to their empirical validity.

Semiparametric and nonparametric approaches are representative of algorithmic modelling culture [Breiman, 2001] that emphasise aspects of learning the complex structure from the available facts and adaptability to the features underlying the data. They are particularly suitable for many hedonic modelling situations, where incomplete knowledge prevents the exact a priori specification of nonlinear or nonstationary components of the functional form. Semiparametric estimators are, more precisely, an intermediate strategy between theory-laden and data-driven estimators that have a restricted learning ability, i.e. semiparametric estimators can approximate functions only within some prespecified classes. Their practical relevance is mainly in balancing the dual goals of low specification error and high efficiency and in enhancing the interpretability of results [Pace, 1995; Anglin & Gencay, 1996]. Nonparametric estimators are by their nature highly flexible and, thus, capable of approximating very general classes of functions (e.g. smooth functions, square integrable functions) that neither require any restrictive, unwarranted prespecification of the functional form of the mean response function nor any specific error distribution assumption. This renders nonparametric estimators to be powerful data-driven tools, albeit highly sensitive to the problem of under-smoothing or overfitting, if local estimation is implemented unduly. Despite the chosen approach, in a model specification dilemma, the modeller has to come up with the relevant set of response and attribute variables, the appropriate functional form between these variables and the adequate error distribution for inference.

Much of the aim of applied hedonic analysis is to produce a reasonable approximation to the generally unknown mean response function. The primary implication of the theoretical literature concerning hedonic prices in the real estate markets is that hedonic relationships are expected to be highly nonlinear due to their locational uniqueness that induces spational heterogeneity of regression surfaces [Wallace, 1996; McMillen & Thorsnes, 2003] that cannot be, in general, specified a priori [Anglin & Gencay, 1996]. Nonlinearity indicates locally changing degrees of curvature in the hedonic function with non-constant characteristic values. Along with the spatial variation, an important source of variability observed in land and hedonic prices is temporal evolution as evidenced by trends and cycles, among others, present in the data. Temporal variation usually induces instability to hedonic functions over time, i.e. data-generating processes are time-varying, which causes series of observations on property prices and attributes to exhibit nonstationarity.

Nonlinearity and nonstationarity of hedonic relationships caused by the spatio-temporal variation are fundamental features that characterise processes in the land markets imposing serious threats to empirical validity of hedonic models that in current practise are pre-dominantly used. The complex question of validity underlying the hedonic model specification can be divided into three sub-problems that involve determining [Pace, 1993 & 1995; Wallace, 1996]:

(1) the relevant set of response and attribute variables;
(2) the appropriate functional form between these variables;
(3) the adequate error distribution for inference.

Nonlinearity and nonstationarity are principal components of spatio-temporal structure, which govern choices in all sub-problems (1)-(3), albeit their effect is most pronounced in the selection of an appropriate functional form. Economic theory and past experience usually provide useful a priori information of what variables should enter the model structure that substantially reduce the threat of omitted variable bias. Phenomena in the land markets are, however, strongly dependent on the particular submarket, time period and land type and, as a consequence, the selection of proper set of dependent and conditioning variables is partially an empirical question, too. For example, in the modelling of land prices, there exists a significant amount of uncertainty concerning the choice of a proper response variable. Should the modeller use the original, total sales prices or the unit prices of a land parcel as the proper response vector and should these be analysed in a logarithmic or untransformed scale? Is it legitimate to apply price levels or should one use the first differences? The effects of time and location are highly important in theory, yet in practice their valid measurement imposes major obstacles and, as a result, no consensus have not yet emerged relating to the determination of their relative influences. The final criterion has to rely on the model’s empirical performance.

Economic theory or previous experience rarely provides any specific, valid guidance on the choice of an appropriate functional form of the hedonic model [Pace, 1993 & 1995; Anglin & Gencay, 1996; Gencay & Yang, 1996]. A prespecified functional form is, however, the fundamental assumption underlying the use of theory-laden parametric models; a poor choice imposes artificial structure on data and significantly invalidates the results of the subsequent analysis[8]. In contrast, nonparametric techniques are data-driven, flexible approaches that can learn much of the genuine structure from available facts and, therefore, allow a greatly reduced attention to the question of which functional form ought to be used.

An ideal statistical method for real estate valuation would possess [Pace, 1993]:

[8] In the parametric modelling context, a common solution to the problem of selecting an appropriate functional form is to consider a set of parametric functions with the objective of finding a model structure that matches the evidence in most measurable respects. However, there is no clear evidence that this practice will be successful in avoiding functional form mis-specification [Anglin & Gencay, 1996; Hannonen, 2005a]. Specification searches can be highly time-consuming and the intrinsic power of these specification tests is somewhat questionable.

(1) low specification error;
(2) robustness against outlying observations;
(3) superior post-sample predictive accuracy;
(4) known statistical properties.

4.2 Management of Nonlinear Behaviour

Local regression techniques can significantly reduce the misspecification error by letting the data to determine the appropriate functional relationship between the response and a set of attributes. Locally weighted regression adapts locally to changing curvature in the hedonic surface by giving more weight to nearby observations [McMillen, 1996] and, therefore, can account for complex nonlinear and certain nonstationary patterns. The local adaptation property, which is achieved by parametric localization [Cleveland & Loader, 1996], makes it a highly attractive tool for estimating spatio-temporally non-homogeneous hedonic functions. Furthermore, any specific assumption underlying the error distribution can be relaxed and, in most cases, derived directly from the evidence e.g. by resampling techniques.

Local regression is a nonparametric method of fitting surfaces to data by smoothing in which parametric functions are locally estimated in the attribute space using weighted least squares [Cleveland & Devlin, 1988; Cleveland & Grosse, 1991; Cleveland & Loader, 1996]. Local regression is most useful when the hedonic function is nonlinear and the form of this nonlinearity is unknown a priori. In land valuation studies the correct functional form is rarely known in advance, although the regression surface is expected to be spatially inhomogeneous with varying degrees of curvature [McDonald & Bowman, 1979; McMillen, 1996; inter alia]. An obvious way to proceed is to use flexible functional forms that can account for complicated spatial, as well as certain temporal, patterns in land values and to infer the form of the hedonic relationship from the available evidence. The main objective in local regression is to smooth the data sufficiently, so that non-systematic variability is ignored, while maintaining as much genuine structure as necessary to produce unbiased results [McMillen, 1996]. This culminates to the well-known bias-variance trade-off: the bias can be reduced only at the expense of variance, and vice versa.

The nonparametric local regression has several benefits that, when combined, make it an attractive tool for modelling complex land value patterns. These include [Hastie & Loader, 1993; Fan & Gijbels, 1995; Fan & Gijbels, 1996; Cleveland & Loader, 1996]:

(1) Attractive minimax properties: the asymptotic minimax efficiency is 100% for commonly used orders among all linear estimators and only marginal loss beyond this class.

(2) Adaptation to various types of designs such as random designs, fixed designs, strongly clustered and nearly uniform designs. It does not require specific smoothness or regulatory conditions.

(3) Easy to understand, interpret and implement. It can utilise the existing statistical theory of least squares regression in a natural way.

(4) Adaptation to boundary bias problem and biases in regions of high curvature.

(5) Adaptation to derivate estimation and spatially inhomogeneous surfaces.

(6) Fast computational methods exist for multivariate smoothing that can also be tailored to different distributional assumptions. Generalises directly to local likelihood estimation framework.

(7) Enables straightforward derivation of response adaptive methods for bandwidth and polynomial order selection.

(8) Has exhibited strong empirical performance in explaining and predicting phenomena in different fields of study.

There are four basic components that have to be specified in practice when using nonparametric local regression. These are [Cleveland & Loader, 1996; Loader, 1999]:

- bandwidth;
- weight function;
- local parametric family;
- fitting criterion.

The bandwidth or smoothing parameter has a critical influence on the performance of local fitting as it significantly controls the model complexity, i.e. the amount of smoothing being performed. Overly small bandwidths lead to undersmoothed estimates that overfit to technical details of the data, while unacceptably large bandwidths oversmooth the data set ignoring relevant systematic features of the dependence of response on the attribute variables. The optimal bandwidth is the one that strikes a balance between bias and variance elements. In real estate valuation practice a natural interpretation can be given to the bandwidth parameter, which sets the number of effective comparables used in the modelling [Pace, 1993].

There are different ways to proceed in specifying the bandwidth parameter: Nearest-neighbour bandwidths are widely applied due to their relative simplicity. In real estate markets typical choices of nearest-neighbour parameters have been from 0.3 to 0.4 [McMillen, 1996; Wallace, 1996; inter alia]. Global cross-validation [Allen, 1974] or generalized cross-validation [Craven and Wahba, 1979] scores, which are motivated by prediction error assessment, are also commonly utilised in bandwidth selection problems. Cross-validation based bandwidth selection rules typically lead to significantly lower values for smoothing parameters [see e.g. Thorsnes & McMillen, 1998] and are often susceptible to the overfitting problem. Indeed, in many instances, the nearest-neighbour bandwidths or global cross validation methods fail to provide a satisfactory fit, and as a result, a different approach to bandwidth selection is often needed.

In the paper 3 “Forecastability of Land Prices: A Local Modelling Approach” [Hannonen, 2006a] locally adaptive bandwidth selection procedures in the spirit of [Clapp, 2003 & 2004] are used to control the model complexity and, in particular, to choose a local smoothing parameter. To derive locally adaptive bandwidths, the local generalized CP criterion is used as a benchmark measure of local goodness of fit. The form of a weight function has much less dramatic influence on the bias-variance dilemma [see e.g. Thorsnes and McMillen, 1998]. Typically, the weight function is chosen to be continuous, symmetric, peaked at 0 and supported on [-1, 1], which assures that the most weight is given to observations close to the fitting point. In the paper 3 the standard tricube function is applied.

The underlying principle of local regression is that the unknown, smooth regression function can be well approximated locally by a member of a parametric class, frequently taken to be polynomials of a certain degree. A high polynomial degree can always provide a less biased approximation to underlying functional relationship but only at the expense of increased variability in the estimate. Therefore, like the bandwidth choice, the degree of the local polynomial affects significantly the bias-variance dilemma and, to some extent, their relative influences are confounded. Asymptotic theory [see e.g. Ruppert & Wand, 1994] suggests that polynomials of odd degree beat those of even degree. Indeed, local linear approximation can almost invariably be preferred to local constant fitting (Nadaraya-Watson kernel estimator), but for local quadratic and cubic fitting the asymptotics perform very poorly [Cleveland & Loader, 1996]. The most natural way to proceed in deciding the appropriate polynomial degree is to utilize available knowledge on global parametric fitting. These considerations [e.g. Hannonen, 2005a] suggest using linear, quadratic or cubic approximations. The exact preference between the different polynomial orders is a priori somewhat uncertain, albeit local quadratic and cubic approximations seem to be preferable as they are expected to reproduce better the peaks and valleys present in the data and, on the other hand, the use of quadratics and cubic polynomials relies on less stringent economic considerations. However, e.g. Clapp [2003] documents that the use of linear approximation resulted a somewhat better out-of-sample MSE than the local constant or quadratic fitting.

Virtually any global fitting criterion can be localized and, therefore, local regression can proceed with the same rich collection of distributional assumptions as in the global case. In the paper 3, error terms are assumed to be independent normal random variables with mean 0 and variance σ^2, which leads conveniently to the least squares criterion. A major problem in practice with the least squares criterion is its high sensitivity to outlying observations, which are very common in studies of land values. In the paper 3 the idea of iterative downweighting [Cleveland, 1979] is used for robustifying the results of local regression by implementing a variant of local M-estimation that uses the input of locally weighted least squares regression. It is easy to implement and usually gives enough protection against extreme observations.

4.3 Results of Local Modelling

The paper 3 "Forecastability of Land Prices: A Local Modelling Approach" [Hannonen, 2006a] deals with the robust nonparametric regression estimation of land prices using local polynomial modelling techniques in the two single localities of the Finnish land market (submarkets of Espoo and Nurmijärvi). In order to reduce sample dependency, the paper examines two data sets of undeveloped sites that are located in different submarkets and associated with partially non-overlapping time frames. To improve theory-dependency of this analysis, the empirical section considers estimation of different localised versions of common econometric models of land prices including dynamic responses. In particular, the paper investigates in depth the predictive accuracy of several different static and dynamic local hedonic models of land prices with adaptive bandwidth selection learning. Static models, whose out-of-sample predictions were analysed, consisted of local double-log, local translog and local cubic expansion models. Dynamic local hedonic models were constructed and their ex-sample forecasting ability was analysed using equilibrium correction mechanisms under two different scenarios: error correction term entered model structure as a linear combination of the original attributes, or as a flexible nonlinear transformation of those characteristics. All local models' post-sample predictions were compared to their proper corresponding global analogues with two data sets differing both in terms of quality and quantity of observations. Furthermore, the results of locally weighted regression were robustified in a straightforward manner by a scheme, which is a variant of M-estimation.

It turned out [Hannonen, 2006a] that the predictive performance of different hedonic models varied significantly across the data sets. In the municipality of Espoo, whose data set can be regarded qualitatively superior (it is manually pre-checked) to the Nurmijärvi case, the use of static local models significantly enhanced the forecasting accuracy as compared to their global alternatives. Specifically, the predictive validity of hedonic models may substantially be improved by local modelling scheme. The results also suggested, beyond any reasonable doubt, that most of the improvement resulted from the increased nonlinearity rather than variable interactions. The most data-congruent local hedonic model, the local cubic expansion model, encompassed the most adequate global alternative, the conventional global double-log model, in most measurable respects in the Espoo case.

The predictive testing provided quite contradictory results in the Nurmijärvi case; this data set contained twice as many observations as the former, yet was more susceptible to any errors due to lack of specific manual pre-checking. In essence, local estimation added no value to the hedonic modelling in terms of the out-of-sample forecasting power. Instead, the most data-congruent specification was either global double-log or translog model, which depended on the measurement of location effect. It appears that the quality of data set is the critical factor, which determinates the utility of flexible estimation techniques in forecasting the land prices, even if the results of in-sample estimation are robustified by some procedure (as here). There also seems to be a genuine association between land prices and housing prices (quarterly price index of single-family houses) in the Nurmijärvi case: the use of the housing price index significantly improved the predictive unbiasedness of any chosen hedonic model. In contrast, for the Espoo case, the use of this index variable would seriously undermine the predictive validity of all model combinations used in the research.

The innovative part of the paper 3, an attempt to unify the concepts of nonlinearity and non-stationary simultaneously into the hedonic model specification by using local regression and error correction mechanisms, utterly failed. This predictive failure is, however, an appearance of potentially a far more profound problem relating to the proper use of equilibrium correction models, and thus to cointegrated processes, at least if they are to be used in the determination of land prices. Over both data sets, the absolute mean prediction errors are marginal, yet their relative effects, when the unit of measurement is taken into account, are unacceptable large (ranging from 23% to infinity!) for any practical land valuation purpose. Furthermore, scaled inner product and gravity indicate that the genuine strength of association between predictions and actual outcomes is low; an important fact that is not revealed simply by looking at the usual correlation coefficient. The most disturbing aspect is, however, that the optimal error correction models clearly pass the conventional specification tests, such as the RESET test and the CUSUM t-test, which measure functional form misspecification and zero forecast innovation mean, respectively. Furthermore, the unit-root t-tests strongly suggest that the observed data is cointegrated, i.e. equilibrium correction model would be a valid description for the data-generating mechanism. This is in direct conflict with the unusual high relative mean prediction errors and low association between model's predictions and observed land prices.

If predictive testing is performed on land markets, the choice of a relevant set of forecasting criteria tends to be critical. It is a sound practice to document several statistics that measure somewhat different aspects of post-sample predictive performance. It is vital to remember to calculate the relative mean prediction errors and, I believe, weighted average-based measures should be preferred over median-based analogues in the evaluation of predictive validity of hedonic models. The strength of association between model's predictions and actual out-of-sample prices should be considered and measures such as scaled inner product, gravity, etc. ought to be preferred to the simple correlation coefficient. In analysing the reliability aspect, the usual root mean squared error and mean absolute prediction error indicators are easily misleading, and the median-based alternatives or mean absolute error indicator would be a better solution.

The main practical problem with local regression is that it gives an infinitive number of different hedonic prices for each regressor, i.e. hedonic prices changes at different data points. Although one can calculate average hedonic prices, there are methodological difficulties in doing that. As a result the main use of local regression is in prediction, not in describing the phenomenon. Thus if the hedonic prices are of main interest, it is advisable to use parametric (or semiparametric) modelling approach. The local modelling is in general susceptible to the overfitting problem and common measures of the goodness-of-fit (such as the coefficient of determination) might easily be misleading. The curse of dimensionality is potentially a major obstacle to any nonparametric method; to avoid this curse one must make sure that the number of estimation sample is large enough and/or the number of explanatory variables is small enough. In the study the number of observations in both samples seemed to be large enough (and the number of regressors small enough) for reliable flexible modelling.

The time periods analysed in the paper 3 are exceptional in that there was a overheating of the whole Finnish economy and/or a subsequent deep depression in those time frames. This makes it difficult to generalise the empirical findings of the paper 3 across time and, as a result, the empirical results should mainly be understood within examined time frames. Also the empirical findings are specific to the analysed submarkets and land type.

5 Evaluation of Hedonic Model's Congruence

5.1 Concept of Congruence and Recursive Least Squares

Congruence is widely used criterion in the evaluation of parametric econometric models; an econometric model is congruent if it is a valid representation of the data-generating process and thus matches the phenomena in all measurable aspects. More stringently, an econometric model is congruent if and only if it possesses [Hendry, 1997b; Bontemps & Mizon, 2001]:

(1) homoskedastic, innovation error terms;
(2) weakly exogenous regressors for the parameters of interest;
(3) constant, invariant parameters of interest;
(4) (at least low-level) theory consistent structures;
(5) data admissible representations with accurate observations.

Each concept corresponds to a different type of information: (1) relates to past data; (2) to present data; (3) to future data; (4) to a priori theory and (5) to measurement system[9]. Further-

[9] Since information is revealed by data, the taxonomy of information sets is designed to classify the data into different groups. Each data group reveals a certain type of information. If we see the data as a single entity, the taxonomy of information is to specify the different aspects of the data. Thus the concept of congruence can be used to the relationship between the data and the model that represents them.

more, if an allowance is made for the information in rival models, this information taxonomy forms the core of econometric model evaluation in the parametric approaches.

The choice of an appropriate functional form of a hedonic model is one of the main methodological concerns in the hedonic estimation process; a poor choice invalidates much of the subsequently analysis. Economic theory of hedonic prices places only a few restrictions on the form of the hedonic function, which therefore cannot generally be specified a priori. Instead, alternative specifications have to be evaluated empirically; the goal is to conclude with a parsimonious model, which has orthogonal regressors and satisfies the necessary conditions for both congruence and encompassing. In the paper 1 “On the Recursive Estimation of Hedonic Prices of Land” [Hannonen, 2005a] is a set of different parametric hedonic models (classical linear model, multiplicative double-log model, additive double-log model, translog model and two standard error correction models) analysed in the Espoo submarket using recursive least squares estimation. The special focus is on the structural stability of estimated hedonic prices of land parcels, i.e. whether constant, invariant parameters of interest can be identified. This is an important issue that is unexpectedly often neglected in applications: Even if the best estimated parametric model would otherwise match the data but still suffer from non-constant parameters (which is a necessary condition for invariance), the model would yield a low level of utility in practical land valuation applications. The problem of non-constant parameters and the issue of invariance in econometric models is the essence of much acclaimed Lukas Critique [Lukas, 1976].

Recursive least squares estimation technique is a useful complement to the ordinary least squares estimation technique in specifying a data-congruent hedonic model and, in particular, for evaluating the structural stability of hedonic prices. Recursive least squares has a number of attractions. Firstly it enables changes in the parameters to be tracked over time and secondly, it produces recursive residuals that are uncorrelated given the classical assumptions. Thirdly once a hedonic model has been estimated from the first k observations, it can easily be updated as each new subsequent observation is added to the data set [Harvey, 1990, pp. 52-53].

Various diagnostic measures can be constructed based on recursive residuals [see e.g. Chow, 1960; Brown et al., 1975; Hansen, 1992] to examine presence (or absence) of a constant parameter model and its invariance. In particular, The CUSUM and CUSUMSQ tests, which are based on the recursive residuals, and their variants offer formal procedures for testing the constancy of hedonic relationships. Recursive estimation also enables a real-time estimation of the necessary hedonic prices and thus could be applied to the construction of a hedonic function in submarkets where only a few observations are available. The philosophy here is, at first, to estimate a general hedonic model at the municipality level and then use these estimation results (the estimated hedonic prices and their covariance structure) as a prior information in a submarket level, where the scarcity of observations otherwise prevents the usual off-line determination of the necessary hedonic coefficients. The modelling approach applied is the so-called general-to-specific, which mimics the theory of reduction with a clear goal of a data-congruent representation that parsimoniously encompasses the information in rival models.

5.2 Results of Recursive Estimation

In the empirical study six different hedonic model structures were estimated by recursive least squares in the submarket of Espoo. Specifically, the parametric hedonic models included:

- linear model
- multiplicative form of double-log model
- additive form of double-log model
- translog model
- linear form of error correction model
- multiplicative double-log form of error correction model.

The empirical investigation indicated that it is possible to construct an overall stable hedonic model, which passes the joint parameter instability test, using an equilibrium correction mechanism, for the multiplicative form of double-log model. This multiplicative double-log form of error correction model also gave the most plausible description among the different model contenders in the model specification phase matching the evidence in most measurable aspects. In particular, in this model the standard coefficient of determination statistic was 0.75, which is over the usual cut-off value of 0.70 used in land valuation and thus indicates that the explanatory power of this model is satisfactory. The multiplicative double-log form of error correction model had the Durbin-Watson statistic of 1.99, which indicates that there exists no first-order autocorrelation in this model structure. Also autoregressive 1-2 test indicated that the model is free of autocorrelation. The two different heteroskedasticity tests indicated that in the optimal error correction model the error terms are homoskedastic. The RESET test was statistically insignificant indicating that there is no serious mis-specification in the multiplicative double-log form of error correction model. This model was also free of any autoregressive conditional heteroskedasticity. However, the CUSUM t-test indicated that a small amount of systematic prediction error was still present in this formulation and the normality test indicated that the residuals were non-normally distributed. Furthermore, the standard error of regression was 0.74, which is high indicating that the model's internal precision might not be as good as desired.

The Hansen's instability test for parcel size variable also implied that the associated hedonic price was somewhat unstable over time. This result produces generalisation problems at least in the case of parcel size variable, i.e. over time it is difficult to predict the influence of lot area variable. In the empirical study with the most data-congruent model, the multiplicative double-log form of error correction model, the most significant attribute variables were identified as, respectively, parcel size (t-value is -20.4), error correction (t-value is -16.5), distance to the Helsinki CBD (t-value is -8.97) and type of a parcel (t-value is -4.08). The house price index variable appeared only as a part of the equilibrium correction term and showed no short-run relation to the dependent unit price variable. Parcel size, distance and parcel type attributes possessed both short-run and long-term impacts on the unit price variable. The empirical study also showed an important aspect of characteristics variables: their statistical significance appears to be strongly time-dependent. Specifically, statistical significance of parcel size, distance, error correction and parcel type variables has increased dramatically over the sample period. The explanation for this phenomenon stems from the nature of t-values: As more data is added the t-values typically increase indicating that the regressors are statistically more significant over time.

The estimated hedonic prices varied somewhat across different model contenders. In the logarithmic or transformed scale the hedonic price of the parcel size variable ranged from -0.79 to -0.0044. The hedonic price of the distance variable ranged from -1.81 to -1.19 in the logarithmic scale. The regression coefficient of the type of a parcel variable was in the range of [-0.37, -0.30] when transformed values were used. In the logarithmic scale the house price index was estimated by a single hedonic price of 2.50. The error correction term was -0.92 in the logarithmic scale. In the original or untransformed scale the hedonic price of the parcel size variable was approximated by a single hedonic price of -0.0015. The hedonic price of the distance variable was in the range of [-1.83, -1.69] in the original scale. The regression coefficient of the house price index variable was estimated by a single value of 0.40. The error correction term was -0.81 in the original scale. The estimated hedonic prices for different model contenders are presented in table 7.

Table 7. Estimated hedonic prices for different model formulations[10].

	Submarket of Espoo				
Variable	Linear model	Multiplicative double-log model	Additive double-log model	Error correction model (linear)	Error correction model (double-log)
Constant	-	-	0.70 [35.7]	-	0.15 [2.94]
Parcel size	-0.0015 [-7.87]	-0.79 [-16.1]	-0.0044 [-11.6]	-0.0015 [-11.2]	-0.77 [-20.4]
Distance	-1.69 [-8.19]	-1.19 [-7.42]	-1.81 [-5.99]	-1.83 [-11.0]	-1.22 [-8.97]
House price index	0.40 [18.9]	2.50 [25.9]	-	-	-
Parcel type	-	-0.37 [-4.74]	-0.30 [-3.07]	-	-0.32 [-4.08]
Time dummy	-4.94 [-2.77]	-	-0.44 [-4.39]	-	-
Error correction	-	-	-	-0.81 [-15.6]	-0.92 [-16.5]
Coefficient of determination	0.41	0.58	NA	0.64	0.75
Number of observations	405	405	405	405	505

t-values are reported in the brackets.

When hedonic models are estimated parametrically, it is sound practise to consider a set of different models instead a single hedonic model. The purpose is to find a data-congruent model structure that matches the evidence best. In this study the fact that a specific error correction model possessed the best fit with respect to several different diagnostic criteria indicates that the land price data is cointegrated.

6 Robust Modelling of Hedonic Prices

6.1. Concept of Robust Modelling

Outlying and influential observations are very common in the land value studies, which may be genuine, faultless values, generated under conditions of some untypical factors or they can contain different errors (such as recording and measurement error; wrong population, etc.). Traditional hedonic modelling techniques, especially the ordinary least squares technique, are very sensitive to outlying observations; even a single outlier can drastically change the results

[10] The translog model reduced in the study to the multiplicative double-log model and thus the estimation results of the translog model are not documented.

and misguide the inferences. In fact, a single sufficiently deviating data point can cause that the least squares estimator breaks down and generates results that are utterly unreliable and uninformative. Robust methods, on the contrary, are not sensitive to outliers or influential observations and, therefore, can tolerate a certain amount of bad observations without the fear that the estimator breaks down and produces completely useless results.

The aim of robust statistics is to investigate the behaviour of estimators, when the basic modelling assumptions (linearity, normality, independence, etc.) are not exactly valid but are at most approximations to reality. To put it slightly differently, the basic aims of robust statistics are [Hampel et al., 1986, p. 11]:

- To describe a structure best fitting the bulk of the data.
- To identify deviating data points (outliers) or deviating substructures for further treatment.
- To identify and give a warning about highly influential data points (leverage points).
- To deal with unspecified serial correlations, or more generally, with deviations from the assumed correlation structures.

In practice the approximate nature of hedonic models is largely result of the occurrence of gross errors, the empirical character of models and only partial validity of theoretical modelling assumptions. In general, the hedonic model can be considered as robust if:

- It is reasonably unbiased and efficient.
- Small deviations from the hedonic model assumptions will not substantially impair the performance of the hedonic model.
- Somewhat larger deviations will not invalidate the hedonic model completely.

The application of robust modelling techniques has been very limited in the real estate markets. The reasons for this are:

- A false belief that a large sample size makes robust modelling tools unnecessary.
- A belief that outlying observations can be discarded by visual inspection or using least squares residuals for this purpose.
- The existence of several different robust modelling tools without a clear instruction which of these techniques is the most appropriate.
- Unfamiliarity with the interpretation of the results of the robust analysis.
- Ignorance of the benefits of the robust analysis in real data sets.

6.2 Results of Robust Modelling

In the paper 5 “On the Robust Modelling of Land Prices” (in Finnish) [Hannonen, 2006c] is investigated the performance of a highly robust estimator in two single localities (Espoo and Nurmijärvi) of the Finnish land market. The results suggest that linear least squares estimates for hedonic prices may be suboptimal, when errors are not normally distributed and, in particular, when errors are heavy-tailed (a common situation in land value studies). There are many different robust estimators, whose common feature is that they give less weight to unusual observations than the ordinary least squares.

In the given paper a very fault tolerant and computationally intensive method, the three-stage MM-estimation, is analysed in the modelling of land prices in the submarkets of Espoo and Nurmijärvi. In the first phase of the MM-estimation is calculated a regression estimate, which is consistent and have a high break-down point, but is not necessarily efficient. In the second phase the scale of errors is estimated, which is based on the residuals of the first phase. In the third phase is calculated the M-estimate of the hedonic prices. The empirical study clearly witnessed that the internal precision, as measured by the standard error of regression statistic, could be improved (circa 25-40%) in the Espoo and the Nurmijärvi case when compared to the ordinary least squares method.

In the empirical study the predictive validity of different estimation approaches were compared to each other. Three different estimation approaches based on the multiplicative double-log model specification were considered, which are:

- the robust three-stage MM-estimation
- ordinary least squares estimation where outliers are not dropped from the model
- ordinary least squares estimation where outliers are dropped from the model.

The post-sample predictive study implied that predictive unbiasedness could be improved significantly with the Espoo case, but not with the Nurmijärvi case, when MM-estimation is used. In the Espoo case all seven predictive validity measures (average prediction error, median prediction error, weighted average error, weighted median error, error centre of gravity, average percentage error and median percentage error) were better with the robust MM-estimation method than approaches based on the least squares estimation. In particular, average prediction error, weighted average error and error centre of gravity are each circa 64 % smaller in the case of robust MM-estimation procedure than with the second best estimation approach: the ordinary least squares estimation procedure where outliers are not dropped from the model. In the Nurmijärvi case, the robust estimation provided contrary results: the ordinary least squares estimation where outliers are not dropped from the model proved to be the most precise approach. In particular, average prediction error, weighted average error and error centre of gravity are each over 60 % smaller in the case of ordinary least squares estimation where outliers are not dropped from the model than in the case of robust MM-estimation.

The estimated hedonic prices of the MM-estimation are presented in table 8; In essence they are slightly different from the usual least squares estimates, and specifically, the statistical effects of parcel size and house price index variables were emphasised in the robust estimation case. In the Espoo case, when MM-estimation is applied, the most significant explanatory variables are, respectively, the house price index variable (t-value is 36.60), the parcel size variable (t-value is -23.40), the distance to the Helsinki CBD (t-value is -10.00) and the parcel type variable (t-value is -4.83). In the Nurmijärvi case, when MM-estimation is used, the most significant regressors are, respectively, the parcel size variable (t-value is -26.71), the house price index variable (t-value is 16.63), the distance to the Helsinki CBD variable (t-value is -7.95), the distance to the parish village of Nurmijärvi variable (t-value is -5.93) and the parcel type variable (t-value is -3.69).

Table 8. Hedonic prices of robust MM-estimation.[11]

	Submarket of Espoo	**Submarket of Nurmijärvi**
Variable	MM-estimation	MM-estimation
Constant	-	3.70 [5.60]
Parcel size	-0.80 [-23.40]	-0.77 [-26.71]
Distance 1	-1.14 [-10.00]	-0.91 [-7.95]
Distance 2	-	-0.23 [-5.93]
House price index	2.48 [36.60]	1.57 [16.63]
Parcel type	-0.26 [-4.83]	-0.14 [-3.69]
Standard error of regression	0.43	0.43
Number of observations	400	793

t-values are reported in the brackets.

The fact that the robust MM-estimation results are fundamentally different between the Espoo and the Nurmijärvi case, indicates that the land price formulation is different in these submarkets and specific to each submarket. Therefore the results of one submarket cannot be generalised to another submarket.

7 Predictive Assessment of Hedonic Models of the Study

Traditionally in the land valuation applications the goodness-of-fit of a hedonic model is assessed by its explanatory power and internal precision. The coefficient of determination statistic (or some modification thereof) measures this explanatory power and it is generally agreed in land valuation that this statistic should be over 0.70 in order for a hedonic model to be satisfactory and adequate. The standard error of regression statistic measures the internal precision aspect and this statistic should be in land valuation under 0.30-0.40 depending on the problem. In an appropriate hedonic model the error terms should typically be independent, normally distributed random variables, whose expected value is zero and variance is constant, which can be tested by different statistical indicators. Furthermore, hedonic prices should be constant. The explanatory variables in a proper hedonic model should be statistically significant, i.e. their t-values should typically be in absolute terms over two and corresponding p-values under 0.05. Multicollinearity should not also exist as evidenced by e.g. low VIF-values. Although these all are important aspects of the fit, they are not always sufficient. In some cases the model contender passes (as in this dissertation in the case of the paper 1 [Hannonen, 2005a]) through these different specification tests, yet the genuine predictive ability of this model might not optimal, or even low. Therefore the post-sample predictive testing should always be considered to be certain how well a hedonic model works in practice, e.g. how good estimates it produces for unknown land prices.

[11] In table 8 are reported the standard errors of regression instead of the coefficient of determinations because the latter cannot be calculated in a usual manner.

In econometrics, therefore, predictive accuracy is perhaps the single most important operational criterion in the evaluation of performance of a chosen model. The success of hedonic model-based forecast depends on [Hendry, 1997a]:

(1) the existence of structure;
(2) whether such structure is informative about the future;
(3) the proposed method capturing the structure;
(4) the exclusion of irregularities that swamp the structure.

The aspects in (1)-(2) are characteristics of the economic system and the last two of the chosen forecasting method. When structure is understood as a systematic relation between the entity to be forecast and the available information, the conditions in (1)-(4) are sufficient for forecastability.

The fundamental idea underlying the hedonic approach is the hedonic hypothesis that implies land prices, which are appearances of the corresponding land values, should be modelled statistically conditional on a small set of fundamental characteristics rather than analysed directly in terms of a large number of similar properties and their unconditional prices. The relevancy of this statistical conditioning is analogous to testing whether a hedonic-based model can predict land prices accurately in terms of some predefined criteria such as mean-squared forecast error. The predictive failure implies that the conditioning is irrelevant and, specifically, unpredictability of land prices, if their conditional mean equals its unconditional mean.

There are numerous different indicators for post-sample predictive assessment of hedonic models [e.g. Case et al., 2004] and the relative ranking of the performance of various models varies according to the applied accuracy measure. However, not all measures can be considered equally informative in the context of hedonic modelling of land prices: of paramount importance are those criteria underlying the mean prediction error and the strength of association between the model's predictions and observed land prices. Variance-based indicators, albeit commonly utilised in studies, are secondary for most practical land valuation purposes. In the papers 2, 3 and 5 predictive power of chosen hedonic models is evaluated by a number of different forecasting criteria.

Mean prediction error is evaluated in this study by (1) arithmetic average prediction error, (2) median prediction error, (3) weighted arithmetic average and median prediction errors, (4) error centre of gravity and (5) by arithmetic average and median percentage errors. If unit land prices are analysed (as in these papers), the weights are directly obtained by using parcel size to give the total mass of errors, which obviates the need of estimating the unknown density function. The use of average prediction errors is often, in land value studies, a more plausible alternative than median-based indicators of central tendency of prediction errors, since latter measures down-weight too heavily the influence of unusual prediction errors as compared to typical prediction errors since they ignore relevant information about distances between errors. Also median-based mean error indicators tend to produce, in many cases, highly similar and indecisive results between different model candidates. The major problem of using arithmetic average prediction error or median prediction error, and their weighted analogues, is that these are absolute measures, which depend on the unit of measurement. Therefore, the relative versions of the basic indicators, i.e. arithmetic average and median percentage errors, are also documented for comparing mean prediction errors between models, where responses are measured on a different scale.

Three measures of strength of the association between predictions and observed out-of-sample land prices are reported. First, the usual correlation coefficient is calculated, which is a useful measure of statistical relation in the case of normally distributed variables and when the focus is on the co-variation of variables. The major problem of using the classical correlation measure in land valuation studies lies in its strong dependency on the normality assumption, which is typically violated by the influence of aberrant observations, whose effect is squared in the denominator, which, in turn, tend to lead to highly similar standard deviations between different model alternatives. Consequently, calculated correlation coefficients tend to produce not only invalid but also indecisive results. The use of the scaled conventional inner product (or some weighted version of it, but this seems to add very little new information) is preferable, which measures the strength of prediction in the direction of actual out-of-sample land prices scaled by the total uncentered variation of land prices. It generates more valid and decisive results that are not strongly dependent on any particular distributional assumptions. Also the gravity [see McMillen, 2001] is reported, which seems to be a viable measure of strength of association; it usually produces results that parallel the use of the scaled inner product.

Root mean squared error is the most commonly used measure of success of numeric prediction, which mainly controls the reliability or variability of predictions, not the actual predictive validity or predictive unbiasedness. Furthermore, this statistic is overly sensitive to outlying observations tending to exaggerate the variance of prediction errors of model choices in which the prediction error is larger than the others (which is typical in land valuation studies). This sensitivity or lack of robustness to extreme errors can also cause the root mean square error to produce results that are indecisive. Mean absolute error is generally a more appropriate indicator of predictive variability, and is especially suitable in cases of outlying prediction errors. Another useful measure of predictive variability is simply the predictive range expressed in terms of minimum and maximum predictions; the shorter the interval, the less variability is expected in predictions. A widely used measure of predictive variability is mean absolute percentage error [see e.g. Makridakis and Hibon, 2000], which is reported here with the associated robust version of it: median absolute percentage error. They have also been criticized for the problems of asymmetry and instability, when the data is small (as here).

In this dissertation one of key findings was that a hedonic model might well satisfy the common ex post statistical criteria (such as high coefficient of determination, low standard error of regression, low p-values, high t-values, low VIF-values, constant parameters, no autocorrelation, no heteroskedasticity, etc.) and yet the model might produce predictions that are not accurate. Specifically, it was concluded on the basis of these ex post good-of-fit measures that a specific error correction model [Hannonen, 2005a] would be the most data-congruent specification, when different hedonic models were assessed parametrically. However, the paper 3 [Hannonen, 2006a] showed that the actual post-sample predictive accuracy of this model was low with very high relative prediction errors.

8 Conclusions

In this dissertation the question of applicability of hedonic pricing methods beyond the ordinary least squares estimator is addressed. The investigated methods in this study included: (a) recursive least squares estimator, (b) structural time series models, (c) local polynomial models, (d) wavelets and (e) robust MM-estimation. The chosen hedonic methodology, when investigating the behaviour of land prices in local markets, should depend on the nature of the problem and, in particular, on what specific aspects of the problem are being emphasised. As a general rule, no single method encompasses the alternatives in all measurable respects and all methods have their strengths and weaknesses. In practice some simplifying assumptions are always necessary to carry out the research and hardly all dimensions of the problem can be addressed within the single approach. A sort of hybrid approach may be needed. The key observation of this study is, however, that there is much scope for methods beyond the ordinary least squares estimator in explaining and predicting the land prices in local markets. This is especially true in the submarket of Espoo, where the use of unconventional methods of the study showed that significant improvements could be achieved in hedonic models' explanatory power and/or predictive validity when the methods of this research are used instead of the orthodox least squares estimator.

In the Espoo case structural time series models, local polynomial regression and robust MM-estimation all generated more precise results in terms of post-sample prediction power than the conventional least squares estimator. Among these methods the local polynomial regression tended to produce the most unbiased results. Furthermore, the recursive estimation and wavelets added new research information on the structural stability of hedonic prices and long-term movements of land prices, respectively. It seems quite strongly on the basis of empirical experimentation that the determination of land prices in the municipality of Nurmijärvi could be best explained by the use of unobserved component models (local level model for trend specification and two different cycle components) among the investigated methods and applying the conventional double-log model specification for regression effects. The flexible local polynomial modelling and three-stage MM-estimation surprisingly added no value in terms of greater post-sample precision in the Nurmijärvi case. A different submarket thus tended to generate different results.

The major practical implication of the results of this study is that the applied unconventional methods could be well used in land valuation applications in order to achieve improved accuracy. The orthodox hedonic method based on the least squares solution is not as precise as the methods or at least some of them investigated in this research. The results also imply that land markets of the study are highly dynamic systems and the determination of land prices is clearly nonlinear.

In the paper 1 recursive least squares estimation was investigated for evaluating structural stability of hedonic prices. The empirical study on undeveloped sites of the submarket of Espoo indicated that it is possible to construct an overall stable hedonic model using an equilibrium correction mechanism, which gave the most plausible description among five different model contenders. In practice this implies that the chosen error correction model could be used for prediction purposes in land value applications. Furthermore, recursive estimation enables a real-time estimation of hedonic prices and thus could be applied to the construction of a hedonic function in submarkets where only a few observations are available. In the empirical study statistically the most significant attribute variables were, respectively: parcel size, error correction term, distance to the Helsinki CBD and type of a parcel.

In the paper 2 spatio-temporal variation of land prices in two single localities (submarkets of Espoo and Nurmijärvi) was analysed by means of structural time series modelling formalism that combines the flexibility of a time series model with that of the interpretation of a regression analysis. The use of unobserved component models resulted to significant improvements in the post-sample predictive accuracy when compared to the results of the orthodox least squares estimator. In predictive testing, for 3 out of 4 different model formulations, the unobserved component approach generated only a marginal average prediction error when compared to the orthodox hedonic models, which, in contrast, yielded to a considerable amount of systematic prediction error. This implies that the structural time series modelling paradigm offers a more viable alternative to the hedonic analysis of land prices than the conventional approach based on least squares estimates.

The paper 3 considers in depth the predictive accuracy of several different static and dynamic robust local hedonic models of land prices with adaptive bandwidth selection learning. All local models' post-sample predictions were compared to their proper corresponding global analogues with two data sets (submarkets of Espoo and Nurmijärvi) differing both in terms of quality and quantity of observations. The article tries to simultaneously combine the concepts of nonlinearity and nonstationary into same hedonic model specification by using local regression and error correction mechanisms. It turned out that local estimation of land prices resulted in a significant improvement in predictive accuracy over the global hedonic models, but only with the second data set, which is the high-quality one (the Espoo case) having been manually pre-checked for errors. The local estimation of equilibrium correction mechanisms resulted in a predictive failure, which is inherently caused by some profound problems with error correction processes, and thus with co-integrated processes, when modelling the temporal dimension of land prices.

The paper 4 analyses the low-frequency temporal variation of unit land prices in two single localities, the municipalities of Espoo and Nurmijärvi, of the Finnish land markets by the use of the wavelet transforms. These transforms are nonparametric orthogonal series estimators, which are capable of providing the necessary time and frequency information of land prices simultaneously in a highly flexible fashion. In the empirical section of this paper both the raw and the quality-adjusted (adjusted using a local regression) unit price series were analysed. Wavelet estimators yielded to meaningful and plausible representations for the low-frequency unit land price fluctuations. The estimated cycles and trends were all nonlinear and, in particular, the behaviour of cyclical component was highly curvelinear and transient in time. The findings strongly suggest that the submarkets in question are not in a steady state, but are continuously evolving in time. It seems that much of the temporal variation present in the untransformed series is, in fact, explained by the quality differences in the attribute variables. The use of quality corrections produced significant improvements in the internal reliability of results.

The paper 5 considers a highly fault-tolerant modelling method, the three-stage MM-estimation method in the econometric analysis of land prices. The method is computational intensive, where in the first phase is calculated a regression estimate, which is consistent and have a high break-down point, but is not necessarily efficient. In the second phase the scale of errors is estimated, which is based on the residuals of the first phase. In the third phase is calculated the M-estimate of the hedonic prices. The MM-estimation method is not sensitive to outliers or influential data points and it can stand a certain amount of bad observations and erroneous data points. The empirical investigation indicated that the internal precision of estimated hedonic models could be improved by 25-40% on both data sets (submarkets of Espoo and Nurmijärvi) when compared to the standard least squares estimator. Furthermore, the post-sample predictive accuracy of the hedonic model could be improved in the Espoo case, when using the MM-estimation.

How well can the results of this dissertation be generalised? It is well known that land prices and their attributes tend to vary across different submarkets, different time period and different land type. This study supported this view. First of all, the estimation results generated different findings in different submarkets (in this dissertation: Espoo and Nurmijärvi). It seems quite strongly that the land price formation process is special to the submarket in question and two different markets tend to produce different empirical findings. Secondly, the time periods of the studies in this dissertation were exceptional as they contained an overheating of the whole Finnish economy and/or a deep depression, both of which reduces the possibility that results could we generalised over time. It should therefore be concluded that the empirical findings should mainly be understood in the context of the investigated time interval and not generalised across time. Thirdly, the land type in this dissertation is specific and represents undeveloped land; other land types would produce differing results. All in all, the empirical findings in this dissertation are unique to specific submarkets, time frames and land type and should not be generalised outside the samples analysed.

Although empirical findings are specific, the methods used in this dissertation are general and could be used in a wide range of different research problems in land markets.

References

Addison, P. (2004). "The Little Wave with a Big Future", *Physics World*, March 2004, 35-39.

Allen, D.M. (1974). "The Relationship between Variable Selection and Data Augmentation and a Method of Prediction", *Technometrics*, 16, 125-127.

Anglin, P.M. & Gencay, R. (1996). "Semiparametric Estimation of a Hedonic Price Function", *Journal of Applied Econometrics*, Vol. 11, No. 6, 633-648.

Asabere, P.K. (1981a). "The Determinants of Land Values in an African City: The Case of Accra, Ghana", *Land Economics*, 57, 385-397.

Asabere, P.K. (1981b). "The Price of Urban Land in a Chiefdom: Empirical Evidence on a Traditional African City, Kumasi", *Journal of Regional Science*, 21, 529-539.

Asabere, P.K. (1991). "Zoning and Industrial Land Values: The Case of Philadelphia", *AREUEA Journal*, Vol. 19, No.2, 154-160.

Asebere, P.K. & Huffman, F.E. (1991). "Historic Districts and Land Values", *Journal of Real Estate Research*, Vol. 6, No. 1, 1-7.

Atac, J. & Margo, R.A. (1998).""Location, Location, Location!"The Price Gradient for Vacant Urban Land: New York, 1835 to 1900", *Journal of Real Estate Finance and Economics*, Vol. 16, No. 2, 151-172.

Barnett, C.J. (1985). "An Application of the Hedonic Price Model to the Perth Residential Land Market, *Economic Record* (Australia), Vol. 61, No. 172, 476-481.

Bontemps, C. & Mizon, G.E. (2001). "Conguence and Encompassing", *Discussion Papers in Economics and Econometrics No. 1017*, Department of Economics, Southampton University.

Brachinger, H.W. (2003). *Statistical Theory of Hedonic Price Indices*, http://www.unifr.ch /stat/forschung/publikationen/Braching-StatHedonic.pdf.

Brandao, A.S.P. & Feder, G. (1995). "Regulatory Policies and Reform: The Case of Land Markets", (ed.) Frischtak, C., *Regulatory Policies and Reform: A Comparative Perspective*.

Breiman, L. (2001). "Statistical Modeling: The Two Cultures", *Statistical Science*, Vol. 16, No. 3, 199-231.

Brown, R., Durbin, J. & Evans, J. (1975). "Techniques for Testing the Constancy of Regression Relationships over Time", *Journal of the Royal Statistical Society*, Series B, 37, pp. 149-172.

Case, B., Clapp, J., Dubin, R. & Rodriquez, M. (2004). "Modelling Spatial and Temporal House Price Patterns: A Comparison of Four Models", *Journal of Real Estate Finance and Economics,* Vol. 29, No. 2, 167-191.

Chaudhry, M.K., Myer, F.C.N., Webb, J.R. (1999). "Stationary and Cointegration in Systems with Real Estate and Financial Assets", *Journal of Real Estate Finance and Economics*, 18:3, 339-349.

Chinoine, D.L. (1981). "Farmland Values at the Urban Fringe: An Analysis of Sales Prices", *Land Economics*, 57, 353-362.

Chow, G.C. (1960). "Tests of Equality Between the Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 591-605.

Clapp, J.M. (2003). "A Semiparametric Method for Valuing Residential Locations: Application to Automated Valuation", *Journal of Real Estate Finance and Economics*, Vol. 27, No. 3, 303-320.

Clapp, J.M. (2004). "A Semiparametric Method of Estimating Local House Price Indices", *Real Estate Economics*, Vol. 32, No. 1, 127-160.

Cleveland, W.S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots", *Journal of the American Statistical Association*,74, 829-836.

Cleveland, W.S. & Devlin, S.J. (1988). "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting", *Journal of the American Statistical Association*, Vol. 83, No. 403, 596-610.

Cleveland, W.S. & Grosse, E. (1991). "Computational Methods for Local Regression", *Statistics and Computing*, Vol. 1, No. 1, 47-62.

Cleveland, W.S. & Loader, C.R. (1996). "Smoothing by Local Regression: Principles and Methods", *Statistical Theory and Computational Aspects of Smoothing*, (eds.) Härdle, W. & Schimek, M.G., Physica-Verlag.

Colwell, P.F. (1990). "Power Lines and Land Value", *Journal of Real Estate Research*, Vol. 5, No. 1, 117-127.

Colwell, P.F. (1998). “A Primer on Piecewise Parabolic Multiple Regression Analysis via Estimations Chicago CBD Land Values”, *Journal of Real Estate Finance and Economics,* Vol. 17, No. 1, 87-97.

Colwell, P.F. & Munneke, H.J. (1997). “The Structure of Urban Land Prices*”, Journal of Urban Economics,* Vol. 41, No. 3, 321-336.

Colwell, P.F. & Munneke, H.J. (1999). “Land Prices and Land Assembly in the CBD”, *Journal of Real Estate Finance and Economics,* Vol. 18, No.2, 163-180.

Colwell, P.F. & Munneke, H.J. (2003). “Estimating a Price Surface for Vacant Land in an Urban Area”, *Land Economics,* Vol. 79, No. 1, 15-28.

Colwell, P.F. & Sirmans, C.F. (1980). “Nonlinear Urban Land Prices”, *Urban Geography,* 1, 141-152.

Craven, P. & Wahba, G. (1979). “Smoothing Noisy Data with Spline Functions”, *Numerische Mathematik,* 31, 377-403.

Crowley, P.M. (2005). “An Intuitive Guide to Wavelets for Economists”, discussion papers, 1/2005, Bank of Finland.

Dowall, D.E. (1993). *The Role and Function of Urban Land Markets in Market Economies*, University of California, http://www.usaid.gov/locations/europe_eurasia/dem_gov/local_gov/pdf_docs/dowall.pdf.

DTZ Pieda Consulting (2002). *Land Values and Implications for Planning Policy*, Scottish Executive Social Research.

Dunford, R.W., Marti, C.E. & Mittelhammer, R.C. (1985). “A Case Study of Rural Land Prices at the Urban Fringe Including Subjective Buyer Expectations”, *Land Economics,* Vol. 61, No. 1, 10-16.

Durbin, J. & Koopman, S.J. (2002). *Time Series Analysis by State Space Methods,* Oxford Statistical Science Series #24, Oxford University Press.

Engle, R.F. and Granger, C.W. (1987). “Co-integration and Error Correction: Representation, Estimation and Testing”, *Econometrica*, 55, 251-276.

Fan, J. & Gijbels, I. (1995). “Data-driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaptation”, *Journal of Royal Statistical Society*, B57, No. 2, 371-394.

Fan, J. & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Monographs on Statistics and Applied Probability, 66, Chapman & Hall.

Gencay, R., Selcuk, F. & Whitcher, B. (2002). *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics,* Academic Press.

Gencay, R. & Yang, X. (1996). “A Forecast Comparison of Residential Housing Prices by Parametric versus Semiparametric Conditional Mean Estimators”, *Economic Letters*, 52, 129-135.

Craig, L.A., Palmquist, R.B. & Weiss, T. (1998). “Transportation Improvements and Land Values in the Antebellum United States: A Hedonic Approach”, *Journal of Real Estate Finance and Economics*, Vol. 16, No.2, 173-189.

Graps, A. (1995). “An Introduction to Wavelets”, *IEEE Computational Science & Engineering*, Summer 1995, 50-61.

Guntermann, K.L. & Thomas, G. (2005). “Parcel size, Location and Commercial Land Values”, *Journal of Real Estate Research*, Vol. 27, No. 3, 343-354.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. & Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley Series in Probability and Mathematical Statistics.

Hannonen, M. (2005a). “On the Recursive Estimation of Hedonic Prices of Land”, *Nordic Journal of Surveying and Real Estate Research*, Vol. 2, No. 2, 30-56.

Hannonen, M. (2005b). “An Analysis of Land Prices: A Structural Time-Series Approach”, *International Journal of Strategic Property Management*, Vol. 9, No. 3, 145-172.

Hannonen, M. (2006a). “Forecastability of Land Prices: A Local Modelling Approach”, *Nordic Journal of Surveying and Real Estate Research*, Vol. 3, No. 1, 20-57.

Hannonen, M. (2006b). “An Analysis of Trends and Cycles Using Wavelet transforms”, *International Journal of Strategic Property Management*, Vol. 10, No.1, 1-21.

Hannonen, M. (2006c). “On the Robust Modelling of Land Prices” (in Finnish), *Maanmittaus*, Vol. 81, No. 1-2, 5-27.

Hansen, B.E. (1992). ”Testing for Parameter Instability in Linear Models”, *Journal of Policy Modelling*, 14, 517-533.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Harvey, A.C. (1990). *The Econometric Analysis of Time Series*, second edition, Philip Allan.

Harvey, A.C. (1997). “Trends, Cycles and Autoregressions”, *The Economic Journal*, 107 (January), 192–201.

Harvey, A.C. & Shephard, N. (1993). “Structural Time Series Models”, (eds.) Maddala, G.S., Rao, C.R. & Vinod, H.D., *Handbook of Statistics*, Vol. 11, Elsevier Science Publishers B.V.

Hastie, T. & Loader, C. (1993). “Local Regression: Automatic Kernel Carpentry”, *Statistical Science*, Vol. 8, No. 2, 120-143.

Heinonen, T. (1993). *The Site Price of Detached, Semi-Detached and Row Houses and Its Development in Finnish Cities in years 1985-1991* (in Finnish), Publication No. 70 of the National Land Survey of Finland, Helsinki.

Hendry, D.F. (1997a). ”The Econometrics of Macroeconomic Forecasting”, *The Economic Journal*, 107 (September), 1330-1357.

Hendry, D.F. (1997b). ”On Congruent Econometric Relations: A Comment”, *Carnegie-Rochester Conference Series on Public Policy*, 47, 163-190.

Hiltunen, A. (1993). *Indicators of the Price Development of Residential Sites* (in Finnish), Publication 2:1993 of the Housing Government, Helsinki.

Hiltunen, A. (2003). *On the Price Formation of Single-Family House Properties: An Analysis of Comparable Sales Prices for Single-Family Houses in Some Regional Areas in Finland* (in Finnish), Doctoral Thesis, Helsinki University of Technology, Espoo.

Isakson, H.R. (1997). ”An Empirical Analysis of the Determinants of the Value of Vacant Land”, *Journal of Real Estate Research*, Vol. 13, No. 2, 103-114.

Isakson, H.R. (2002). ”The Linear Algebra of the Sales Comparison Approach”, *Journal of Real Estate Research*, 24, 117-128.

Johnson, M.S. & Ragas, W.R. (1987). “CBD Land Values and Multiple Externalities”, *Land Economics*, Vol. 63, No. 4, 337-347.

Kanerva, V. (1974). *A Research on Residential and Commercial Sites and the Influencing Factors in Certain Population Centres of the County of Uusimaa* (in Finnish), Liceniate’s Thesis, Helsinki University of Technology, Espoo.

Kantola, J. (1967). *A Research on the Price Formation of Residential and Commercial Sites in Certain Population Centres of the County of Uusimaa* (in Finnish), Liceniate’s Thesis, Helsinki University of Technology, Espoo.

Koopman, S. J., Harvey, A.C., Doornik, J.A. & Shephard, N. (1999). *Structural Time Series Analysis, Modelling and Prediction using Stamp*, Timberlake Consultants, London.

La Croix, S.J. & Rose, L.A. (1989). “Urban Land Price: The extraordinary Case of Honolulu, Hawaii”, Urban Studies, 26, 301-314.

Lehmusto, M., Kanto, A. & Sorsa, A. (1991). *The Local Price Development of Residential Sites in the Helsinki Metropolitan Area in 1980’s* (in Finnish), Publication B 1991:1 of the YTV, Helsinki.

Leväinen, K.I. (1991). *A Calculation Method for a Site Price Index*, the Association of Finnish Cities, Helsinki.

Leväinen, K. & Virkkunen, H. (1987). *Sufficiency and Price Development of Residential Bulding Land in the Finnish Cities in 1980's* (in Finnish), Publication No. 358 of the Association of Finnish Cities, Helsinki.

Lin, T.C. & Evans, A.W. (2000). "The Relationship between the Price of Land and Size of Plot When Plots Are Small", *Land Economics*, Vol. 76, No. 3, 386-394.

Loader, C. (1999). *Local Regression and Likelihood*. Springer Series in Statistics and Computing, Springer-Verlag.

Loader, C. (2004). "Smoothing: Local Regression Techniques", *Handbook of Computational Statistics*, (eds.) Gentle, J., Härdle, W. & Mori, Y., Springer-Verlag.

Lucas, R.E. (1976). Econometric Policy Evaluation: A Critique, in: R.E. Lukas (ed.), *Studies in Business-Cycle Theory*, Cambridge (MA), MIT Press: 104-130.

Lukkarinen, P. (1981). *The Prices of Residential Sites in Finland* (in Finnish), Publication of the State Technical Research Centre, Espoo.

MacKay, D.J.C. (1992). *Bayesian Methods for Adaptive Models*, PhD thesis, California Institute of Technology.

Mackenzie, D. (2001). "Wavelets; Seeing the Forest and the Trees", *Beyond Discovery*, National Academy of Sciences, 1-8, http://www.beyonddiscovery.org/includes/DBFile.asp?ID=958.

Makridakis, S. & Hibon, M. (2000). "The M3-Competition: Results, Conclusions and Implications", *International Journal of Forecasting*, 16, 451-476.

Manning, C.A. (1988), "The Determinants of Intercity Home Building Site Price Differences", *Land Economics*, Vol. 64, No. 1, 1-14.

McDonald, J.F. & Bowman, H.W. (1979). "Land Value Functions: A Re-evaluation", *Journal of Urban Economics*, Vol. 6, No. 1, 25-41.

McDonald, J.F. & McMillen, D.P. (1998). "Land Values, Land Use and the First Chicago Zoning Ordinance", *Journal of Real Estate Finance and Economics*, Vol. 16, No.2, 135-150.

McMillen, D.P. (1996). "One Hundred Fifty Years of Land Values in Chicago: A Nonparametric Approach", *Journal of Urban Economics*, Vol. 40, No. 1, 100-124.

McMillen, D.P. (2001). "Nonparametric Employment Subcenter Identification", *Journal of Urban Economics*, Vol. 50, No. 3, 448-473.

McMillen, D.P. & Thorsnes, P. (2003). "The Aroma of Tahoma: Time-Varying Average Derivates and the Effect of a Superfund Site on House Prices", *Journal of Business and Economics Statistics*, Vol. 21., No. 2, 237-246.

Miyamoto, K., Paez, A. & Uchida, T. (2001). "Spatial Association and Heterogeneity Issues in Land Price Models", *Urban Studies*, Vol. 38, No. 9, 1493-1508.

Myhrberg, O. (1971). *Influences of Construction Political Injunctions and Restrictions on Single-Family and Coastal Settlement* (in Finnish), Publication of the Institute of Real Estate Studies No. A1/1971, Helsinki.

Pace, R.K. (1993). ”Nonparametric Methods with Applications to Hedonic Models”, *Journal of Real Estate Finance and Economics,* Vol. 7, No. 3, 195-204.

Pace, R.K. (1995). ”Parametric, Semiparametric, and Nonparametric Estimation of Characteristic Values within Mass Assessment and Hedonic Pricing Models”, *Journal of Real Estate Finance and Economics*, Vol. 11, No.3, 195-217.

Peltomaa, H. (2001). *The Depreciation of Real Estate Caused by Expropriation and Its Compensability* (in Finnish), Doctoral Thesis, Edita Inc., Helsinki.

Polikar, R. (2001). *The Wavelet Tutorial,* http://users.rowan.edu/~polikar/WAVELETS/ WTtutorial.html.

Potepan, M.J. (1996). “Explaining Intermetropolitan Variation in Housing Prices, Rents and Land Prices, *Real Estate Economics,* Vol. 24, No. 2, 219-245.

Raittinen, P. (1975). A Research on the Price Formation of Sites in Certain Population Centres of the County of Uusimaa (in Finnish), Thesis Required for a Diploma, Helsinki University of Technology, Espoo.

Ramsey, J.B. (2000). “The Contribution of Wavelets to the Analysis of Economic and Financial Data”, Silverman, B. & Vassilicos, J. (eds.), *Wavelets: The Key to Intermittent Information*, OUP, 221-236.

Ramsey, J.B. (2002). “Wavelets in Economics and Finance: Past and Future”, *Studies in Nonlinear Dynamics and Econometrics,* Vol. 6, No. 3, 1-27.

Rosen, S. (1974). “Hedonic Price and Implicit Markets: Product Differentiation in Pure Competition”, *Journal of Political Economy,* Vol. 82, Issue 1, pp. 34-55.

Ruppert, D. & Wand, M.P. (1994). “Multivariate Locally Weighted Least Squares Regression”, *The Annals of Statistics*, Vol. 22, No. 3, 1346-1370.

Schleicher, C. (2002). “An Introduction to Wavelets for Economists”, working paper 3/2002, Band of Canada.

Schulz, R.M.A. (2003). *Valuation of Properties and Economic Models of Real Estate Markets*, Dissertation, Berlin.

Serra, M.V., Dowall, D.E., Motta, D. & Donovan, M. (2004). *Urban Land Market and Urban Land Development: An Examination of Three Brazilian Cities: Brasilia, Curitiba and Recife,* working paper 3/2004.

Shonkwiler, J.S. & Reynolds, J.E. (1986).”A Note on the Use of Hedonic Price Models in the Analysis of Land Prices at the Urban Fringe”, *Land Economics*, Vol. 62, No. 1, 58-63.

Thorsnes, P. & McMillen, D.P. (1998). “Land Value and Parcel Size: A Semiparametric Analysis”, *Journal of Real Estate Economics and Finance*, Vol. 17, No. 3, 233-244.

Trivelli, C. (1997). *Agricultural Land Prices*, http://www.fao.org/sd/LTdirect/LTan0016.htm.

Tsoodle, L.J. (2006). “Factors Influencing Kansas Agricultural Farm Land Values”, *Land Economics*, Vol. 82, No. 1, 124-139.

Tulaca, S.A., Myer, F.C.N. and Webb, J.R. (2000). “Dynamics of Private and Public Real Estate Markets”, *Journal of Real Estate Finance and Economics*, 21:3, 279-296.

Vidakovic, B. (1999). *Statistical Modelling by Wavelets*, John Wiley & Sons Inc.

Virtanen, P. (2000). *Municipality’s Land Policy* (in Finnish), Rakennustieto Inc., Tampere.

Wallace, N.E. (1996). ”Hedonic-Based Price Indexes for Housing: Theory, Estimation and Index Construction”, *Economic Review*, No. 3, 34-48.

Yamazaki, R. (2000), “Empirical Testing of Real Option Pricing Models using Land Price Index in Japan”, *Journal of Property Investment and Finance*, Vol. 19, No. 1, 53-72.

Änkö, O. (1969). *The Price Level of Land in the Region of Helsinki in 1967* (in Finnish), Publication of the State Technical Research Centre No. 143, Helsinki.

RESEARCH PAPERS

[1] “On the Recursive Estimation of Hedonic Prices of Land”, *Nordic Journal of Surveying and Real Estate Research*, Vol. 2, No. 2, 30-56.

[2] “An Analysis of Land Prices: A Structural Time-Series Approach”, *International Journal of Strategic Property Management*, Vol. 9, No. 3, 145-172.

[3] “Forecastability of Land Prices: A Local Modelling Approach”, *Nordic Journal of Surveying and Real Estate Research*, Vol. 3, No. 1, 20-57.

[4] “An Analysis of Trends and Cycles Using Wavelet transforms”, *International Journal of Strategic Property Management,* Vol. 10, No.1, 1-21.

[5] “On the Robust Modelling of Land Prices” (in Finnish), *Maanmittaus*, Vol. 81, No. 1-2, 5-27.